INTERNATIONAL JOURNAL OF BEHAVIORAL DEVELOPMENT, 1998, *22* (2), 403–422

False Belief Understanding in Young Children: Explanations do not Develop Before Predictions

Heinz Wimmer and Heinz Mayringer
University of Salzburg, Austria

Two studies contrasted children's ability to predict a wrong action with their ability to explain such an action in a standard unexpected transfer task. It was found that the majority of 3½- to 4½-year-old children was unable to explain in an appropriate way why the protagonist looked for the critical object in the wrong place and, therefore, exhibited at least as much difficulty with explanation as with prediction. This finding speaks against Fodor's (1992) critical account of the standard false belief tasks. According to Fodor, these tasks induce children to rely on too simple action prediction heuristics ("Predict that the agent will act in a way that will satisfy his desire") although they possess an understanding of belief and desire as joint causes of action. Analysis of children's inadequate explanatory attempts showed that in the majority of these answers they referred to the protagonist's desire to get the object or to the actual location of the object. These desire and reality orientations in explanation are similar to response tendencies in prediction and suggest a lacking in understanding of the causal links between misleading informational conditions, epistemic states, and resulting actions in younger children.

The traditional view in "children's theory of mind" research has it that around 4 years of age children become able to attribute false beliefs to other persons and to themselves and this is taken by many developmentalists as indication that the child begins to understand the mind as a representational medium (e.g. Perner, 1991; Wellman, 1990). The age claim is based on findings from variants of the so-called false belief tasks. In Wimmer and Perner's (1983) unexpected transfer task, the child has to predict where the doll, Maxi, will look for his chocolate when coming back from the playground, after the mother—in Maxi's absence—had

Requests for reprints should be sent to Dr Heinz Wimmer, Hellbrunnerstrasse 34, A-5020 Salzburg, Austria; e-mail: heinz.wimmer@sbg.ac.at.

Thanks are due to Natalia Hofer and Sabine Hofman for their help with testing the children in this study. Josef Perner helped by critically reading a first version of the manuscript. We appreciate the thoughtful comments of several reviewers.

unexpectedly transferred the chocolate from the green cupboard, where Maxi had put it, to the blue cupboard. Similarly, in Hogrefe, Wimmer, and Perner's (1986; also Perner, Leekam, & Wimmer, 1987) deceptive box task, the test question was what the other child (waiting outside) will answer when asked about the content of a matchbox, after the subject himself/herself had found out that, for example, a toy car was in the matchbox. Obviously, both tasks are prediction tasks: From the misleading information the other person is exposed to, the child has to infer the other person's false belief about the critical state of affairs; and from the false belief, the child has to predict what the other person will do or will say. Three-year-old children tend to fail these tasks and to respond with the actual location of the critical object in the unexpected transfer task and with the actual content in the deceptive box task. In contrast, 4-year-olds tend to give belief-based predictions. Similar age trends were observed when the false belief of the other person was stated directly and children had only to predict the resulting action (Moses & Flavell, 1990). The shift around 4 years was also observed when children had to recall a false belief of their own (Astington & Gopnik, 1988; Wimmer & Hartl, 1991; Riggs & Robinson, 1995), and around this age the appearance-reality distinction gets acquired as documented by Flavell and collaborators (e.g. Flavell, Flavell, & Green, 1983).

The nontraditional view claims that children considerably younger than 4 years do understand false belief and that the standard false belief tasks do underestimate young children's true understanding (for a review, see Flavell & Miller, 1997). It was proposed that 3-year-olds in the unexpected transfer task misunderstood the prediction question "Where will he look?" as "Where should he look?" or as "Where will he look eventually?" (Siegal & Beattie, 1991), or that they get distracted from the agent's belief by the salience of reality in the false belief task (e.g. Mitchell & Lacohee, 1991). The most cogently argued critique of the standard false belief tasks was provided by Fodor (1992), who, in defence of a nativist epistemological position, proposed that 3-year-olds in the standard false belief tasks are misled to rely on too simple prediction heuristics. For example, in the unexpected transfer task, 3-year-olds rely on the heuristic (H1) "Predict that the agent will act in a way that will satisfy his desire". They rely on this heuristic because of limited processing capacity that prevents them from routinely checking an agent's belief (in addition to desire) for deriving an action prediction. The specific further argument is that the standard unexpected transfer task misleads children into trusting their simple heuristic because the heuristic applied to the task runs smoothly and gives a unique prediction—for example that the agent will look for the chocolate where it actually is. However, these young children, according to Fodor, share the basic assumption of common sense theory of mind, namely, that actions are caused by desire and belief. This understanding would become evident in

appropriately structured test situations. One of Fodor's proposals was to change the standard false belief task in a way so that H1 would not result in a unique action prediction, for example, by splitting up the chocolate and transferring it to not one but two different locations in Maxi's absence. However, the expected improvement of 3-year-olds' action predictions in this altered version of the standard task was not found (Wimmer & Weichbold, 1994). Another implication of Fodor's theoretical account of why 3-year-olds fail in the standard tasks is that they should not fail in a false belief explanation task. Obviously, if children do have a general theoretical understanding that actions are based jointly on desire and belief, then—given an action which is done in fulfilment of a given desire but blatantly contradicts this desire (e.g. Maxi going to the empty cupboard in order to get his chocolate)—they should have little difficulty to explain such a misguided action. Fodor considered this implication of his theory as already proven by Bartsch and Wellman (1989), who reported that 3-year-olds have much lesser difficulty to explain a wrong action in terms of false belief than to predict such an action. This finding not only supports Fodor's specific critique of the standard false belief tasks but also the other already mentioned critical accounts which stressed misunderstanding of the test question or distraction from the agent's belief by the salience of reality. Obviously, an explanation question such as "Why does Maxi go to this (empty) cupboard to get his chocolate?", may be less prone to misunderstanding than "Where does Maxi go to get his chocolate?" which may invite a misunderstanding as "Where should he go . . .?" Distraction by reality should also be reduced because physical realisation of the wrong action (Maxi moving to the wrong place) should counter the distraction exerted by the actual location of the critical object.

Given the importance of the explanation data for the critical evaluation of the traditional view on theory of mind acquisition, it is surprising that only two published studies have directly focused on the contrast between children's belief-based action predictions versus their belief-based action explanations. Furthermore, there are concerns about the replicability and the validity of the findings. A closer look at the procedure and the findings might be useful. Bartsch and Wellman (1989) in their seminal study used stories like: "Look, here's Bill. Bill has a cut, see? And he wants a Band-Aid". Then Bill (a puppet) was moved to a Band-Aid box, which before was shown to be empty. The explanation question was: "Why do you think he's looking in there?" followed by the prompt: "What does Bill think?" In the prediction tasks the beginning of the stories was analogous to the explanation tasks, but for asking the prediction question "Where will Bill look for Band-Aid?" always two boxes were introduced: For example, a Band-Aid box without Band-Aid and a plain box containing Band-Aid. With this procedure, 3-year-olds were found to show 66% correct

explanations but only 31% correct predictions. One critical aspect of the explanation task is that the box, which the protagonist is approaching, is empty. Therefore, when the child should explain why Bill is going to look into the empty Band-Aid box, there is no plausible alternative to the correct answer which is suggested by both the appearance of the box and by the protagonist's action. In contrast, for the prediction task the wrong alternative (the plain box containing Band-Aid) is the salient answer to the question: "Where will he look for Band-Aids?" Even the "easy" answers to the explanation question must have been infrequent, because the majority of the 66% correct responses were *not* given to the explanation question but to the following prompt: "What does he think?" Further doubts about the validity and replicability of the Bartsch and Wellman finding have been raised by Moses and Flavell (1990, pp. 941f) and Perner (1991, pp. 309f).

However, a recent study by Robinson and Mitchell (1995) with a new procedure, consistent with Bartsch and Wellman (1989), also found the explanation-over-prediction advantage among 3-year-olds. Robinson and Mitchell used a clever procedure which in the explanation version involved two identically clothed twin puppets. Both twins were presented as watching when a ball was hidden in one of two cupboards. In the absence of one of the twins the unexpected transfer occurred, after which the second twin also left. For the explanation question both twins reappeared with the intention to find the ball, one moving to the empty cupboard, the other to the cupboard with the ball. The explanation question then was asked for the twin standing besides the empty cupboard: "So, this one's gone to the wrong place, hasn't he. Why's he gone to the wrong place, is it because he went outside, or because he stayed inside?" In the alternative prediction task, the twins were dressed differently to make obvious which one was absent or present during the unexpected transfer, and the prediction question was asked for the twin who was absent during transfer of the ball: "Now, where will he go first of all to look for the ball, here or here?" In several experiments, Robinson and Mitchell found an advantage of explanation over prediction. For example, in Experiment 1 with older 3- and 4-year-olds the percentages were 85% correct for explanation, but only 37% correct for prediction. However, as noted by Perner (1995, p. 253) this difference is less impressive because without any belief understanding there should already be a sizeable advantage for explanation. Obviously, for a child without any understanding of false belief, the two answer alternatives for the explanation question ("Was the twin at the wrong location outside or not?") are of equal plausibility, whereas in the prediction task the actual location is the obvious but wrong choice of the two alternatives for such a child. Therefore, those children who predicted correctly would also be correct in the explanation task, and the remaining children who predicted incorrectly may have resorted to guessing in the explanation task. Given about 40% correct

predictions, this implies expected 70% correct explanations compared to observed 85%. Furthermore, some children in their response to the explanation question may have relied on a simple association between "gone to the wrong place" and "was outside". More seriously, replicability of the explanation advantage seems to be a problem. According to one of the authors, in further work with the twins procedure, 3-year-olds were not found to differ reliably from guessing level in the explanation version of the task (Robinson, personal communication).

Although not directly focused on the contrast between prediction and explanation, there are observations that, counter to Bartsch and Wellman (1989) and Robinson and Mitchell (1995), suggest that explanation of a misguided action in terms of false belief may be as difficult as prediction (Clements & Perner, 1994; Wimmer & Hartl, 1991; Wimmer & Weichbold, 1994; Yoon & Yoon, 1994). For example, Wimmer and Weichbold (1994) used the standard unexpected transfer task and if a child had predicted correctly that Maxi will look into the empty box, the child was asked to justify the prediction (e.g. "Why will Maxi look for his chocolate in this [empty] cupboard?"). When the child had predicted incorrectly, the experimenter supplied the correct prediction and asked for justification ("No, Maxi is going to the green cupboard to get his chocolate. Why will Maxi look for the chocolate in the green cupboard?"). Wimmer and Weichbold found a close-to-perfect association between prediction and explanation among their 3- and 4-year-old children: With three exceptions all children who correctly predicted Maxi's misguided action (mostly 4-year-olds) were able to explain this action, and not a single child with incorrect prediction (mostly 3-year-olds) was able to explain the following experimenter-provided misguided action. The problem with Wimmer and Weichbold's procedure is that children may have been confused when they themselves made a wrong prediction and were then corrected by the experimenter. This may have prevented them from explaining the experimenter-provided misguided action.

In summary, this review of findings suggests that it may be premature to take for granted that an understanding of false belief emerges first in young children's attempts to explain actions and only later in children's attempts to predict actions. Similarly, it may be premature to take for demonstrated that the traditional false belief prediction tasks underestimate young children's true competence with belief-based reasoning. The present research takes up the issue again. Basically, we repeated the Wimmer and Weichbold (1994) prediction/explanation procedure which, as already noted, found developmental consistency between prediction and explanation. However, to account for the mentioned problem with this procedure we added a straightforward explanation condition, where the wrong action of the protagonist was immediately given and the child was asked why the actor

went to the wrong location to get the desired object. If the prediction/explanation condition in fact confuses the child, then a marked improvement in children's explanation should be observed in the explanation condition.

The main question, however, is how children's explanations are related to their predictions. If the critical accounts of the standard false beliefs tasks are valid, then one would expect that young children should have fewer difficulties with explanation (at least in the explanation condition) than with prediction. As already noted, this follows from Fodor's (1992) theory that the traditional unexpected transfer prediction task misleads young children to rely on the too simple heuristic: "Predict that the agent will act in a way that will satisfy his desire", despite their understanding of desire and belief as joint causes of action. Similarly, explanation should be easier than prediction when the mentioned misunderstandings of the prediction questions occur (i.e. "Where will he look?" misunderstood as "Where should he look?"). Also, the explanation advantage would follow, if it is the case that in the prediction task children are simply distracted from belief-based reasoning by the salience of reality. On the other hand, if the standard false belief prediction task provides a rather valid assessment of children's emerging understanding of false belief, then one would expect no explanation advantage but consistency of prediction and explanation performance.

STUDY 1

Method

Participants. We tested 60 children from three kindergartens, which mainly serve a working class and lower-middle class neighbourhood in the city of Salzburg, Austria. The sample was divided into three age groups with 20 children (about the same number of boys and girls) in each group. The youngest group ranged from 3;4 years to 4;6 years (mean: 3;11), the middle group from 4;7 years to 5;4 years (mean: 4;11), and the oldest group from 5;5 years to 6;3 years (mean: 5;11).

Design and Procedure. Each child was confronted with two different stories (book story, ice-cream story), one in the explanation condition, the other in the prediction/explanation condition. Assignment of stories to conditions and order of conditions were both counterbalanced within each age group. Both stories were acted out with Playmobil figures in model scenarios. The model for the book story were two rooms on a 35cm × 25cm platform. A back wall (on a long side of the platform) and a wall separating the platform in two halves depicted the rooms. The left room (from the

child's position) was the cloakroom (with a mini-clothes hanger), the right room was the playroom (with a table). In both rooms there was a cupboard (green matchbox in the cloakroom, red matchbox in the playroom). The book story was about Ann (girl Playmobil figure) who hides her preferred picture book in the green cloakroom cupboard, before she goes with the other children to the playground (disappears behind the back wall). In her absence, the kindergarten teacher (adult Playmobil figure) finds the book in the cloakroom and puts it back into the book cupboard in the playroom. After a while Ann comes back from the playground and wants to get her book. In the *explanation condition* the story continued: She goes directly in the cloakroom to get her book (acted out with the Ann figure). The *explanation question* then was: "Why then does Ann go to the cloakroom to get her book?" (German: "Warum geht die Anna denn in die Garderobe, um sich ihr Buch zu holen?"). In the intonation of the question, stress was placed on "cloakroom" to mark that this was the fact that should be explained. Furthermore, by including Ann's goal ("to get the book") into the question, children should be prevented from explaining the action via Ann's desire to have the book again. If children still answered with Ann's intention (e.g. with "Because she wants to look at the book"), then the experimenter confirmed this answer and repeated the explanation question (e.g. "Yes, Ann wants to look at the book. But why then does she go to the cloakroom [stressed] to get the book?"). When a child did not respond to the action explanation question at all or produced an inadequate answer, the experimenter rephrased the question as a belief explanation question: "Ann goes into the cloakroom, because she thinks the book is in this cupboard (pointing). Why then does Ann think the book is in this cupboard?"

In the *prediction/explanation condition* Ann's action when coming back from the playground and wanting her book was not presented by the experimenter, but was asked for. At this point in the story, the Ann figure was placed before the wall separating the two rooms and the child was asked the *prediction question*: "Where then does Ann now go to get her book?" (German: "Wo geht die Anna denn jetzt hin, um sich ihr Buch zu holen?"). In the case of no response two answer alternatives were provided: "Does Ann go to the cupboard in the cloakroom (pointing) or does she go to the cupboard in the playroom (pointing)?" In the case of a correct answer, the explanation question(s) followed in the same way as in the explanation task. When the child responded to the prediction question in "realist" manner with "playroom cupboard" or by pointing to this cupboard, the experimenter provided the correct prediction: "No, look, where Ann is going. She is going to the cloakroom cupboard", before asking the explanation question(s). The final question in the book story for both tasks was the *control question*: "Remember, Ann has put the book in this (cloakroom) cupboard. Why is the book no longer in this cupboard?"

The ice-cream story was structurally analogous to the book story, but superficial similarities (e.g. two cupboards as critical locations) were avoided. The story was enacted in the model of a village set up with Playmobil houses and other material on a 105cm × 65cm base. In front of the child was the house of a boy named "Peter" (boy Playmobil figure). Two streets left from Peter's house, one diagonally to the left to the playground (with trees and Playmobil playground equipment) and one diagonally to the right to the train station (toy station building). The playground was connected with the train station by a third street. The story was about Peter, who initially meets the ice-cream man (Playmobil figure with ice-cream cart) in front of the playground. Peter wants to buy ice-cream, but has no money with him. He decides to get ice-cream in the afternoon and goes home (Peter figure disappears in the house). While Peter eats his lunch, the ice-cream man goes to the train station to sell ice-cream there (the ice-cream man is moved from the playground to the station). After lunch Peter wants to have ice-cream and gets his money. For the explanation condition, the story continues with Peter starting to go to the playground (Peter figure is placed on the street to the playground). The questioning for the explanation condition was then analogous to the book story beginning with: "Why then does Peter go to the playground to buy an ice-cream?" For the prediction/explanation condition, the experimenter put the Peter figure in front of Peter's house where the two streets split and asked: "Where then does Peter now go to, to buy an ice-cream?", in the case of no response, followed by: "Does he go to the playground or does he go to the train station?" The control question for the ice-cream story was "In the morning Peter met the ice-cream man in front of the playground. Why couldn't he buy an ice-cream in the morning?"

Results

Number of Correct Explanations. Children's responses to the explanation questions of a story were scored as correct when a child produced an appropriate response to the action explanation question (39 cases) or to the follow-up belief explanation question (15 cases). Acceptable answers either specified relevant epistemic states of the actor (e.g. "Because she doesn't know where the book is now") or mentioned the earlier location of the critical object (e.g. "Because the book was in here"). (A detailed analysis of correct and incorrect responses is given later). Mentioning a relevant story fact counted as a correct answer to the control question. For example, to "Why is the book no longer in the cupboard?" we counted as correct: "Because the teacher put the book back into the book cupboard" or "Because the book is in the book cupboard now".

An initial finding was that children were not negatively affected in their explanatory attempts when they themselves attempted to predict the actor's action and often were corrected by the experimenter (in the prediction/explanation condition) than when the misguided action was immediately provided by the experimenter in the explanation condition. The percentage of 45% correct responses in the prediction/explanation condition, which has been criticised as being potentially confusing, was only slightly lower than the percentage of 50% in the explanation condition: Binomial test, $N = 15$, $x = 6$, $P > .60$. There was also no reliable difference in the number of correct explanations for the first versus for the second story (50% vs. 45% correct: Binomial test, $N = 15$, $x = 6$, $P > .60$). Only story content made a difference with the book story leading to more correct explanations (57% correct) than the ice-cream story (38% correct: Binomial test, $N = 15$, $x = 2$, $P = .007$). The advantage of the book story may have to do with the fact that the "earlier location" of the critical object was physically dominant as the empty cloakroom cupboard in the book story, whereas in the ice-cream story there was no such dominant reminder of the ice-cream man's earlier position although, of course, the playground (his earlier location) was still present. This difference may be important because the majority of the explanations did not refer to the actor's false belief but to the "earlier location" (see the analysis of the explanations later). Table 1 shows the number of children in each age group with two, one, or no correct explanations.

From Table 1 it is evident that the majority of the youngest age group failed both explanation questions, whereas among the two older age groups only a minority of children exhibited such a failure. The difference in the number of children with at least one explanation correct was reliable between the two younger age groups [$\chi^2(1, N = 40) = 6.40$, $P = .01$], whereas the difference between the two older age groups was not [$\chi^2(1, N = 40) < 1$]. There was no reliable difference between boys and girls [$\chi^2(1, N = 60) < 1$]. The difficulty in explaining a misguided action for the youngest age group was not due to a general difficulty with understanding or answering "why-questions" about story facts. Of the 14 3½- to 4½-year-old children

TABLE 1
Number of Children with Two, One, or No Explanations Correct

	Age (Years)		
Explanations	3½–4½	4½–5½	5½–6½
2	3	7	11
1	3	7	5
0	14	6	4

who for both stories failed to explain the misguided action, 9 were able to pass the control questions for both stories, and the remaining 5 subjects passed at least one of the control questions. There was not a single child in any age group who failed to answer both control questions. Furthermore, as shown later, the majority of incorrect answers to the explanation question were related to the question in a relevant way, indicating some understanding of the question.

Table 2 shows the relationship between prediction and explanation performance. As children were asked only one prediction question and because in the case of explanation no guessing was possible, we credited a child with explanatory competence when at least one explanation question was answered correctly.

Table 2 shows that explanation was not easier than prediction. From the percentage totals (upper section of Table 2) it is obvious that explanation was of about the same difficulty as prediction. For the youngest age group, prediction tended to be slightly easier than explanation. However, this advantage of prediction was not reliable: Binomial test, $N = 10, x = 3, P > .30$. Overall, the contingencies (lower section of Table 2) show that the consistency of prediction and explanation performance was moderately high (67%).

The comparison between prediction and explanation, shown in Table 2, rests on the assumption that children's correct predictions (e.g. that the actor will look for the book in the "empty" cupboard) were not based on guessing between the two possible locations, but were derived from the actor's belief. Against guessing speaks the fact that the two locations are not of equal plausibility. The actual location (e.g. the cupboard with the book in it) is the obvious answer for a child who does not yet understand false belief and its role in the causation of action. Another possibility for false positive

TABLE 2
Percentage and Number of Children with Correct Answers to Prediction and Explanation Questions

	Age (Years)			
	3½–4½	*4½–5½*	*5½–6½*	*Total*
Independent questions (%)				
Prediction	50	65	75	63
Explanation	30	70	80	60
Contingency between prediction and explanation (*N*)				
Both	3	11	13	27
Prediction only	7	2	2	11
Explanation only	3	3	3	9
Neither	7	4	2	13

prediction performance has to do with the fact that in the present design half of the children had already experienced the explanation condition before they served in the prediction condition. These children might have simply imitated what the experimenter had demonstrated in the preceding explanation condition—namely, that the actor searches in the wrong location without understanding why. Because there were actually somewhat more correct predictions when the prediction condition was presented after the explanation condition than when it was presented before (67% vs. 58%), one could argue that our first comparison is biased in favour of prediction over explanation. However, a second comparison—limited to only those children who had received the prediction condition first—gave results similar to the one including all children. In this second comparison, 58% of the children showed correct prediction and exactly the same percentage had at least one correct explanation. Among the youngest age group ($3\frac{1}{2}$–$4\frac{1}{2}$ years), 40% showed correct prediction and the same percentage exhibited at least one correct explanation.

The most conservative assessment of prediction performance would be to count as correct only when the child responded to the prediction question with the empty location and subsequently, in response to the explanation question, justified this answer appropriately (e.g. with "Because Ann had put the book in here"). Obviously, this assessment would guard against guessing and also against imitation of the experimenter-provided action. Twenty-two of the 38 correct predictions were correctly justified and 10 of the 16 instances of correct prediction without justification occurred among children who had received the prediction condition after the explanation condition. Overall, 37% of the sample showed correctly justified action prediction, whereas 50% of the children explained correctly the experimenter-provided action in the explanation condition. This explanation advantage was not reliable [McNemar's test, $\chi^2(1, N = 60) = 2.3$, $P > .10$]. For the youngest age group the difference between prediction and explanation was small, with 15% of the children showing a correctly justified prediction and 25% showing a correct explanation of the demonstrated action.

In summary, none of the three prediction versus explanation comparisons suggests that explanation of a misguided action in terms of false belief or misleading information conditions is easier and develops earlier than the prediction of such an action.

Content of Explanations. About half (i.e. 54%) of the correct answers to the action explanation question referred to the earlier location of the book or the ice-cream man and the other half (i.e. 46%) included an epistemic term. The following answers are typical for both types of correct answers.

Earlier location answers:
Experimenter: "Why then does Ann go to the cloakroom to get her book?"
Child: "Because Ann had put the book in here."
"Because the book was in here."

Epistemic answers:
Experimenter: "Why then does Peter go to the playground to buy an ice-cream?"
Child: "Because he thinks the ice-cream man stays there." (8 instances)
"Because the ice-cream man was there and he didn't know that he is now near the station." (6 instances)
"Because he saw the ice-cream man there." (4 instances)

The preponderance of the earlier location answers was even more marked in children's correct answers to the belief explanation question (e.g. "Ann goes to the cloakroom, because she thinks the book is in this cupboard [pointing]. *Why then does Ann think the book is in this cupboard*?"). This belief explanation question was asked when children failed to answer correctly the preceding action explanation question. Among the 18 correct responses to the belief explanation questions only 4 answers referred to ignorance or lack of perceptual access (e.g. "Because she doesn't know that it is in the book cupboard now"). All the other responses which were considered correct referred to the "earlier location".

Among the incorrect responses to the explanation questions many answers referred to the desire of the protagonist as exemplified by the following answers.

Experimenter: "Why then does Ann go to the cloakroom cupboard to get her book?"
Child: "Because she wants her book."
"Because she likes the book."
Ice-cream Story: "Because it (the ice-cream) tastes so good."

Obviously, these answers explain why Ann is going somewhere to get her book and not why she is going to get it from the "empty" location. There were 39 (i.e. 33%) such answers to the first formulation of the action explanation question. It should be remembered that we wanted to discourage children from giving such answers by already stating Ann's goal to get the book in the question so that only the specific action is in need of explanation. After a desire answer to the first formulation of the action explanation question the question was repeated with explicit acknowledgment of the desire (e.g. "Yes, Ann wants to have the book. But

why then does she go to the cloakroom [stressed] to get the book?"). This further attempt to prevent children from desire answers was not particularly successful, because only in 8 instances did a child switch from a desire answer to a correct answer, leaving still 31 instances (26%) of desire answers to the second formulation of the action explanation question. Twelve instances of desire answers (i.e. 19% of incorrect answers) were also given in response to the belief explanation question. In this case, the desire answers are inappropriate ("Why does Ann think the book is in this cupboard?"—"Because she likes to look through her book").

Another type of inappropriate response to the action explanation questions could be termed "reality answers", because they referred to the present location of the critical object. These answers constituted 28% of all incorrect responses. A typical dialogue was like this:

Experimenter: "Why then does Ann go to the cloakroom cupboard to get her book?"
Child: "Because it's in here" (points to playroom cupboard).
"Because it's no longer in here" (points to cloakroom cupboard).
"Because the teacher has put the book away."

These answers sound absurd, because they would have better served as explanation for the alternative, not presented action of going to the "new" location and not as explanation for the actually presented action of going to the "old" location. A substantial number of reality answers (18 instances, i.e. 29% of incorrect answers) also occurred in response to the belief explanation question (e.g. "Why then does Ann think the book is in this [cloakroom] cupboard?").

There were also a number of refusals or responses which could not be classified: 23% in case of the action explanation question, 30% in case of the belief explanation question.

STUDY 2

The aim of this study was to examine the generality of the findings of Study 1 with a different sample of children. In particular, we were interested in whether the difficulties of 3- and young 4-year-old children with the explanation of misguided action could be replicated.

Method

Participants and Procedure. The 24 children of this study came from a kindergarten in Salzburg, which in contrast to the kindergartens of Study 1, serves mainly an area with upper-middle class families. The 12 younger

children ranged from 3;8 years to 4;7 years (mean: 4;1), the 12 older children ranged from 4;8 years to 5;10 years (mean: 5;3). Within each age group boys and girls were of about equal frequency. Design, procedure, and scoring of explanations were the same as for Study 1.

Results

Again, as in Study 1, there was no difference at all between children's correct explanations in the explanation condition and in the potentially confusing prediction/explanation condition (46% correct in each). Also, the age trend for correct explanations was similar to the one of Study 1. Among the younger age group, only 3 children (25%) exhibited at least one correct explanation, whereas among the older group, 10 children (83%) did so [$\chi^2(1, N = 24) = 8.2$, $P < .01$]. Of the 12 children in the younger age group, 9 children failed both explanations task, but only 1 failed the control questions for both stories, 4 children answered correctly both control questions, and a further 4 were correct on one control question.

Explanation turned out to be slightly more difficult than prediction. Percentages of children with correct prediction were 50% and 100% for the two age groups, respectively, compared to 25% and 83% for explanation. Combined over the two age groups, there were 5 children with correct prediction but without a single correct explanation, whereas no child showed the opposite pattern: Binomial test, $N = 5$, $x = 0$, $P = .06$.

Inspection of children's answers to the action explanation question showed tendencies similar to Study 1. Only 6 of the 20 correct answers referred to epistemic states and the rest were of the "earlier location" type. Among the 28 wrong answers there were 15 desire answers, 4 reality answers, 6 refusals, and 3 not-classifiable answers.

In summary, Study 2 confirmed the finding of Study 1 that the majority of children in the age range of 3½ to 4½ years are unable to explain a misguided action in an appropriate way. For the children in Study 2, it can be ruled out that low social family background could be responsible for the difficulties.

DISCUSSION

A main finding of the present study is that 3- and 4-year-old children did not find explanation of a misguided action in terms of false belief or misleading circumstances easier than the prediction of such an action. Indeed, a tendency to the opposite was observed. In particular, the large majority of the youngest group (3-year-olds and young 4-year-olds) was unable to explain a misguided action appropriately. A similar difficulty with action explanation was found before by Wimmer and Weichbold (1994). The

present finding, however, stands in marked contrast to Bartsch and Wellman (1989) and Robinson and Mitchell (1995), who had found that 3-year-olds have much less difficulty with explanation than with prediction. The present findings support the critical evaluation of these studies—in particular the difficulty of replication—given in the introduction. The high difficulty of explanation for the 3- and young 4-year-old children constitutes counter-evidence to Fodor's (1992) theoretical account of the standard false belief prediction tasks. As already noted, Fodor argued that 3-year-old children understand belief and desire as joint causes of action, but because of limited processing capacity rely on the simple heuristic: "Predict that the agent will act in a way that will satisfy his desire". The standard unexpected transfer task misleads children into relying on this heuristic because the task is structured in such a way that the simple heuristic gives a unique prediction. Fodor explicitly assumed that—in contrast to prediction (in the standard tasks)—explanation of a misguided action should pose little difficulty. The present finding shows that this implication of Fodor's theory is wrong. Another implication of Fodor's theory—namely, that young children's prediction performance should improve when the simple heuristic does not result in a unique predicition—had also found no support (Wimmer & Weichbold, 1994). The difficulty of explanation, found here for the 3- and young 4-year-olds, also speaks against other attempts to explain away the difficulty of young children in the standard false belief tasks as being due to superficial task factors associated with the action prediction requirement. It is certainly possible that the prediction question: "Where does Ann go now to get her book?" may be misunderstood as: "Where should Ann go now to get her book?" (Siegal & Beattie, 1991) and that the salience of the present location of the critical object may distract children from belief-based action prediction (Mitchell & Lacohee, 1991). The important question, however, is whether these tendencies to respond with the actual than with the believed location are limited to children, who have *no* understanding of false belief, or whether these tendencies prevent belief-based action prediction also for children, who *do* have an understanding of false belief. Our findings speak for the first alternative. In the explanation condition, a misunderstanding of the explanation question similar to the one mentioned for the prediction question is not possible and the salience of the actual object location should be offset by the salience of the actual action of going to the wrong place. Nevertheless, the youngest group found explanation at least as difficult as prediction. Therefore, we would readily admit that in the standard false belief prediction tasks there are biases towards wrong prediction. However, these biases seem to affect predominantly children who have not yet developed a good understanding of false belief. In parenthesis we note that the prediction task, where it is enough to point to the "empty" location, may, under certain circumstances, also lead to false positive answers. For

example, in the present study some children may have imitated without understanding in the prediction/explanation condition what the experimenter had demonstrated in the preceding story, namely, that the main figure goes to the empty location.

Of course, it is difficult to exclude that young children's true ability to explain misguided actions may have been underestimated in the present research. However, we note that nearly all children who consistently failed to explain the misguided actions were able to answer control questions where another story fact had to be explained. The further observation that the majority of the incorrect answers were related to the explanation question in a sensible way (see later) also shows that the question format was not a major problem. Furthermore, difficulties of story comprehension and of memory for story facts are unlikely causes for young children's inability to explain the misguided action. Note that for such an explanation in terms of a false belief no reference to any story fact is necessary. This is due to the fact that the explanation question specifies both the wrong action and the goal, in the service of which the action was performed, for example: "Why then does Ann go to the cloakroom cupboard to get her book?" From this information alone it follows that Ann must act on the belief: "She thinks the book is in the cloakroom cupboard". It should be remembered that after the frequent desire answers to the explanation question several reformulations of the explanation question were used to guide children to a correct answer, but with little success. It is also important to note that no complex syntactic constructions were needed for a correct answer to the explanation question and, in fact, many answers were very simple and only referred to the earlier location of the critical object.

A second main finding of the present research is given by the content of children's wrong answers to the explanation question. As noted, quite often children answered with the protagonist's desire to have or use the critical object and insisted on this answer despite our various attempts to prevent them. On the one hand, these answers show that children superficially understood the explanation question. The desire to have or use an object is an appropriate answer, when one is asked why somebody is looking for a particular object. But obviously the answer misses the point because the question was not simply why the actor was looking for the critical object, but why he/she was looking for the object in the wrong place (e.g. "Why then does Ann go to this [empty] cupboard to get her book?"). However, if there is no understanding of false belief or misleading circumstances as cause of the misguided action, then to explain with desire is the best one can do. In this respect, it is important to remember that even when the desire was acknowledged and the wrong place stressed in the second formulation of the explanation question ("Yes, Ann wants to get her book, but why then does she go to the cloakroom cupboard to get it?"), the majority of the

children with an original desire answer did continue to maintain the desire answer.

Interestingly, the frequent desire answers to the explanation question fit well with the *spirit* of Fodor's (1992) assumption that young children's too simple heuristic for action prediction focuses on desire and neglects belief: "Predict that the agent will act in a way that will satisfy his desire". The only problem is that Fodor had assumed that the young child who relies on this heuristic in the standard false belief prediction tasks, nevertheless understands the role of belief in the causation of action. This is contradicted by the present finding that even in explanation, children responded solely with desire and not with belief. Therefore, the desire-based prediction heuristic may reflect the same underlying problem as the desire answers in the explanation task, namely, young children's inability to understand false belief. The desire answers also fit well with Wellman's (1990) developmental proposal that the child quite early—in the second year—acquires an understanding of desire as mental cause of action, before in a later stage he/she acquires an understanding of belief as additional cause of action.

The second rather frequent type of incorrect answers to the explanation question were references to the actual location ("Why then does Ann go to the cloakroom cupboard to get her book?"—"Because it's in here [playroom cupboard]"). These reality answers, as the desire answers, are also superficially related to the question, because they acknowledge the fact that the question has to do with where the critical object is looked for. Of course, they constitute no explanation, but if there is no understanding of belief or misleading circumstances, then there is little choice. What the child may want to say when he/she responds with the actual location, may be something like this: "I don't know why she is going to the empty location, but she should go to where the book really is". This *misunderstanding* of the explanation question is similar to the one noted for the prediction question, when "Where does she go to get the book?" is misunderstood as "Where should she go to get the book?" Again, as in the case of the desire answers, one can assume that the underlying cause for the misunderstandings of the explanation and the prediction question is the young child's inability to understand false belief.

This discussion of children's wrong answers to the explanation question should be complemented by a discussion of the answers which were counted as indicating an understanding of false belief. This is all the more important as the term *think* (German: *denken*, *glauben*), which is the most direct reference to false belief, occurred in only few of the answers. Such answers may have been infrequent because they are not particularly informative. To respond to "Why then does Ann go to the cloakroom cupboard to get the book?" with "Because she thinks it's there", is more or less a reformulation of the action and might invite the further question "But why does she think

it's there?" Indeed, as noted in the Results section, more than half of the answers which were considered correct were of the *earlier location* type (e.g. "Because it was there before"). This was also the most frequent answer given to the belief explanation question ("Why then does Ann think the book is in this cupboard?"), which was asked when the child did not respond correctly to the action explanation question. Quite similar observations about children's correct explanations of misguided actions or false beliefs have been made by Wimmer and Weichbold (1994) and Wimmer and Hartl (1991). With these answers children obviously refer to the concrete starting point of the causal chain which—together with the protagonist's presence and later absence—eventually leads to the wrong action. The false belief is only the mental intermediary between the objective misleading circumstances and the objective wrong action.

The observation that children explain misguided actions and false beliefs in terms of concrete misleading circumstances is broadly consistent with the theoretical position advanced by Wimmer, Hogrefe, and Sodian (1988). The main hypothesis was that around the age of 4 years children acquire an understanding of informational conditions (e.g. what did the protagonist see . . . , what didn't she see . . .), and this is a critical step in children's ability to attribute epistemic states like knowledge and belief to other persons. In the first narrow formulation of this position it was, for example, argued that 3-year-old children may fail false belief prediction tasks because they do not understand the importance of misleading information conditions and therefore do not infer from such conditions the resulting false belief. It turned out that the difficulty of young children was not limited to the inferential use of informational conditions. For example, young children were found to have the same difficulty when they themselves had actually experienced a false belief and therefore had not to infer it, but had only to identify it as false belief (Astington & Gopnik, 1988; Wimmer & Hartl, 1991). In a modification of the original proposal, Wimmer and Weichbold (1994, also Wimmer, 1993), therefore, suggested that the very concept of belief can only be formed, when the child understands belief causation. According to this view a false belief is not simply a false thought, which was taken to be true (in one's own case) or is taken to be true (in the other's case). False belief, in this view, is a false thought, that was or is fully determined by concrete misleading circumstances. These phenomenal (i.e. taken to be reality) and causal characteristics (i.e. resulting from concrete circumstances) distinguish a false belief from a false guess. That the present children in their successful attempts to explain a misguided action so frequently referred to the concrete misleading situation and so infrequently to a false thought, seems consistent with this perspective. This perspective based on the network view of concept formation sees the ontogenesis of epistemic concepts embedded in the emergence of a broader causal

theoretical framework, which links informational circumstances, inner mental states, and actions.

Manuscript received June 1997
Revised manuscript received July 1997

REFERENCES

Astington, J.W., & Gopnik, A. (1988). Knowing you've changed your mind: Children's understanding of representational change. In J.W. Astington, P.L. Harris, & D.R. Olson (Eds.), *Developing theories of mind* (pp. 193–206). Cambridge, UK: Cambridge University Press.

Bartsch, K., & Wellman, H.M. (1989). Young children's attribution of action to beliefs and desires. *Child Development*, *60*, 946–964.

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

Flavell, J.H., Flavell, E.R., & Green, F.L. (1983). Development of the appearance-reality distinction. *Cognitive Psychology*, *15*, 95–120.

Flavell, J.H., & Miller, P.H. (1997). Social cognition. In D. Kuhn & R.S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language development* (5th ed.). New York: Wiley.

Fodor, J.A. (1992). A theory of the child's theory of mind. *Cognition*, *44*, 283–296.

Hogrefe, J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, *57*, 567–582.

Mitchell, P., & Lacohee, H. (1991). Children's early understanding of false belief. *Cognition*, *39*, 107–127.

Moses, L.J., & Flavell, J.H. (1990). Inferring false beliefs from actions and reactions. *Child Development*, *61*, 929–945.

Perner, J. (1995). The many faces of belief: Reflections on Fodor's and the child's theory of mind. *Cognition*, *57*, 241–269.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J., Leekam, S.R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*, 125–137.

Riggs, K.J., & Robinson, E.J. (1995). What people say and what they think: Children's judgements of false belief in relation to their recall of false messages. *British Journal of Developmental Psychology*, *13*, 271–284.

Robinson, E.J., & Mitchell, P. (1995). Masking of children's early understanding of the representational mind: Backwards explanation versus prediction. *Child Development*, *66*, 1022–1039.

Siegal, M., & 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.

Wimmer, H. (1993). *A revised version of how understanding of informational causation affects children's understanding of false belief.* Unpublished manuscript, University of Salzburg, Austria.

Wimmer, H., & 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., Hogrefe, J., & Sodian, B. (1988). A second stage in children's conception of mental life: Understanding informational accesses as origins of knowledge and belief. In J.W. Astington, P.L. Harris, & D.R. Olson (Eds.), *Developing theories of mind* (pp. 173–192). Cambridge, UK: Cambridge University Press.

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

Wimmer, H., & Weichbold, V. (1994). Children's theory of mind: Fodor's heuristics examined. *Cognition*, *53*, 45–57.

Yoon, M.G., & Yoon, M. (1994). *Correct prediction of intentional action develops earlier than attribution of false belief in three-year-olds.* Unpublished manuscript, Department of Psychology, Dalhousie University, Halifax, Canada.