Acting and Planning on the Basis of a False Belief: Its Effects on 3-Year-Old Children’s Reasoning About Their Own False Beliefs

Cristina M. Atance and Daniela K. O’Neill
University of Waterloo

This study examined 3-year-olds’ explanations for actions of theirs that were premised on a false belief. In Experiment 1, children stated what they thought was inside a crayon box. After stating “crayons,” they went to retrieve some paper to draw on. Children were then shown that the box contained candles and were asked to (a) state their initial belief and (b) explain their action of getting paper. Children who were unable to retrieve their false belief were unable to correctly explain their action. Experiments 2 and 3 ruled out several alternative interpretations for these findings. In Experiment 4, children planned and acted on their false belief. Again, children who were unable to retrieve their false belief were unable to correctly explain their action.

A 3-year-old is shown a crayon box and is asked to state what she thinks is inside. After she has stated “crayons,” the box is opened to reveal candles. Once the box is closed up again, she is asked what she initially thought was inside. Her response, like that of most typical 3-year-olds, is “candles,” not “crayons.” However, consider what would happen if this procedure were modified in the following way: After the child has stated that she thinks the box contains crayons, she is asked to get a piece of paper to draw on. Upon returning with the paper, she is shown the candles. The box is closed up again and she is asked to state what she initially believed it to contain. If she responds “candles,” she is asked to explain why she went to get the paper.

The ability to make sense of our own actions, as well as those of others, constitutes an important aspect of our folk psychology. Yet how, or whether, young children make sense of actions of theirs that were premised on a false belief (i.e., a false-belief-based action) is an area of study that is conspicuously absent from the literature. If we reconsider the example presented in the first paragraph, several scenarios are possible: (a) Asking a 3-year-old child about her own prior false-belief-based action may provide the impetus to allow her to appeal to this belief to explain her action, or (b) the 3-year-old child may not appeal to a false belief—but then how will she explain her action, which only makes sense in the context of having believed something that turned out to be false (e.g., that the crayon box actually contains candles)?

In general, we know more about how children explain the actions of others than about how they explain their own. By 3 years of age, children can appeal to such mental states as desire and (true) belief to explain the actions of other people (e.g., Bartusch & Wellman, 1989). However, whether 3-year-olds can explain the action of another that was premised on a false belief is debatable. Bartusch and Wellman (1989) found that when 3-year-old children were told, “Here’s Jane. Jane is looking for her kitten. The kitten is hiding under the chair. But Jane is looking under the piano,” they were able to appeal to a false belief to explain Jane’s misguided behavior. Sixty-five percent of the 3-year-olds gave at least one explanation that made reference to a false belief (e.g., “She thinks it’s under the piano”).

However, in contrast, more recent research has revealed striking limitations in 3-year-olds’ explanations of behavior premised on a false belief (Moses & Flavell, 1990; Wimmer & Mayringer, 1998; Wimmer & Weichbold, 1994). For example, Moses and Flavell (1990, Experiment 2) presented 3-year-old children with a scenario in which a character with a hurt hand found a Band-Aid box, only to discover that it contained a toy car. When the children were asked to explain why the character was searching in the Band-Aid box, only 3% of their explanations invoked the character’s false belief as the cause of her action (e.g., “Because she thinks there are Band-Aids in there”). Instead, children either made reference to the character’s desire (e.g., “Because she wants Band-Aids”) or to the outcome of the situation (e.g., “Because there’s a car in there”) or provided no response at all. Similar findings have been reported by Wimmer and Weichbold (1994) and Wimmer and Mayringer.

---

Cristina M. Atance and Daniela K. O’Neill, Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

This article is based on the doctoral dissertation of Cristina M. Atance under the direction of Daniela K. O’Neill at the University of Waterloo and on postdoctoral work conducted at the University of Washington Center for Mind, Brain, and Learning. The dissertation work was supported by a doctoral grant from the Natural Sciences and Engineering Research Council of Canada and an Ontario Graduate Scholarship awarded to Cristina M. Atance and by an Operating Grant from the Natural Sciences and Engineering Research Council of Canada to Daniela K. O’Neill. The postdoctoral work was supported in part by the Talaris Research Institute and by an Operating Grant from the Natural Sciences and Engineering Research Council of Canada and an Ontario Graduate Scholarship awarded to Cristina M. Atance.

We thank Rebecca Shultis and Ksenia Kosobutsky for their help with the coding of the data, Rechelle Brooks and Jessica Sommerville for insightful comments on an earlier version of the article, and the children and the parents who volunteered their participation.

Correspondence concerning this article should be addressed to Cristina M. Atance, University of Ottawa, 120 University, Room 407, Ottawa, Ontario K1N 6N5, Canada. E-mail: atance@uottawa.ca
shown that there is a marked improvement between 3 and 4 years (Watson, 2001) of children’s performance on false belief tasks have although the results of a recent meta-analysis (Wellman, Cross, & zoff, 1994; Slaughter & Gopnik, 1996), there exist no studies that improve their false belief reasoning (Gopnik, Slaughter, & Meltzoff, 1997; Gopnik & Wellman, 1992), it is just this sort of conundrum that should provide children with especially salient counterevidence, thus forcing them to appeal to a false belief. Moreover, this inconsistency should be most salient when it concerns actions that are motivated and performed by the self. Although several studies have shown that, over several training sessions, providing children with counterevidence can improve their false belief reasoning (Gopnik, Slaughter, & Meltzoff, 1994; Slaughter & Gopnik, 1996), there exist no studies that have examined how children react to such an “acute” dose of counterevidence.

Addressing how young children explain their own false-belief-based action is important for several reasons. First, children must come to recognize that there are many instances in which their own behavior can only be explained as a function of a false belief. For example, my (false) belief that there is ice cream at home may lead me to stop at the grocery store to buy chocolate sauce. However, when I arrive home and discover that my belief was in fact false (i.e., there is no ice cream), I can nevertheless acknowledge that I bought the chocolate sauce because “I thought that there was ice cream at home.” Thus, I can make sense of my behavior by invoking my false belief. Although such a situation may appear to be a trivial one, it is not difficult to imagine ones that are not. For instance, explaining to my boss that my informing a client of a confidential fact (one that the client was not meant to know) was the result of my (false) belief that the client was aware of this fact, rather than the result of a malicious act on my behalf, is an example of one such situation. Second, the ability to acknowledge a false belief as an explanatory construct constitutes an important learning device. That is, by acknowledging that our false beliefs can lie at the root of our actions, we are then in a position to reflect on these actions and subsequently learn from them. Thus, the next time I am tempted to discuss confidential information, I may make greater efforts to ensure that the person with whom I intend to discuss this information is in fact knowledgeable about it.

To recap, no research has examined how children explain their own false-belief-based actions despite the fact that addressing this issue would contribute to an understanding of an important facet of young children’s reasoning about false belief. Thus, the following “action” paradigm (referred to here as the false-belief-action task) was developed from the unexpected contents task (e.g., Gopnik &...
Astington, 1988). Children were shown a crayon box and were asked to state what they thought was inside. After they stated “crayons,” the experimenter pointed out to the children that there was a piece of paper on the floor and suggested that they go get it to draw on with the crayons. When the children returned with the piece of paper, it was revealed that the box contained candles. Once the box was closed up, the children were asked what they initially believed it to contain. All children (i.e., both those who correctly and those who incorrectly stated their initial belief) were then asked to explain their action. In addition to this crayon box trial, children were given a second false-belief-action trial that was conceptually identical to the first but that involved a juice box that unexpectedly contained sand (juice box trial). Prior to discovering the true contents of the juice box, children were asked to get a cup so that they could drink some juice.

Our main goal in Experiment 1 was to determine how young children would explain their own false-belief-based action. A secondary goal was to directly compare children’s performance on the false belief question with their performance on the action explanation question. More generally, children’s performance on the false-belief-action task speaks to their conceptual understanding of false belief. That is, should children succeed in this type of explanatory context, then one could argue that their false belief reasoning can be enhanced by highlighting a prior action of theirs that was motivated by a false belief. If, on the other hand, children perform poorly in this context, then one could argue that they are not yet aware that false belief can be an important determinant of their own behavior.

Experiment 1

Method

Participants

Participants were 63 children (30 boys and 33 girls; age range = 3 years 1 month to 3 years 11 months; mean age = 3 years 6 months). Fourteen additional children were excluded from this final sample for refusing to state their initial belief about the contents of the box (n = 4), for being unable to complete the procedure owing to fussiness (n = 4), or because of procedural difficulty (n = 6; see Juice box trial section below). Children were predominantly from White, middle-class families and were recruited through advertisements in local malls, day-care centers, and other public centers.

Materials

Stimuli included a crayon box that contained candles, an orange juice box that contained sand, a sheet of 8½ × 11 in. white paper, and a green plastic cup.

Design and Procedure

Children were tested individually in a laboratory playroom. All testing sessions in this and all subsequent experiments were videotaped. Prior to the testing session, a piece of paper and a cup were placed alongside a wall adjacent to the table where the child and the experimenter would be seated during the session. The order of the two false-belief-action trials (crayon box and juice box) was counterbalanced. The procedure for these two trials was as follows:

Crayon box trial. Children were shown a crayon box and were asked to state what they thought was inside. Once children had stated that they believed the box contained crayons, the experimenter pointed out that there was a piece of paper on the floor and suggested that they go get it to draw on with the crayons: “Look, there’s some paper over there. Why don’t you get it to draw on with the crayons?” After children had gone to get the paper and returned to the table, they were shown that the crayon box actually contained candles.

Juice box trial. Children were shown a juice box and were asked to state what they thought was inside. Once children had stated that they believed the box contained juice, the experimenter pointed out that there was a cup on the floor and suggested that they go get it so that they could drink some juice: “Look, there’s a cup over there. Why don’t you get it so that you can drink some juice?” After children had gotten the cup and returned to the table, they were shown that the juice box actually contained sand. In 6 cases (listed in the Participants section as being eliminated because of procedural difficulties), children unexpectedly stated that they did not want to go get the cup because they were not thirsty. This occurred equally often on trials in which the juice box was presented first and on trials in which it was presented second. These children were not included in the final sample (N = 63), as it was imperative to the design that children act on their false belief.

False belief test and reality control questions. For both trials, once children had seen the unexpected contents of the box, and the box had been closed up, they were asked the following false belief question: “Before, when you first saw the box all closed up like this, what did you think was inside?” If children incorrectly responded “candles [sand],” they were immediately asked the following action explanation question: “Why did you go get the paper [cup] then?” For those children who correctly responded “crayons [juice],” an ensuing reality control question was asked: “What is inside the box?” This question was asked to ensure that children were indeed differentiating between their false belief about the contents of the box and the true contents of the box. These children were then asked the following modified action explanation question: “Why did you go get the paper [cup]?” The wording of this question was changed slightly for this group of children for pragmatic reasons. That is, these children were not being confronted with the mismatch between their stated belief (e.g., “candles”) and their prior action (e.g., going to get some paper) but were merely being asked to recall why they had performed their action. Thus, pragmatically, it did not make sense to include the word “then” at the end of the question. In essence, the explanations of this group of children provided a comparison with the explanations of the group of children who passed the false belief question.

Coding

Children were considered to have passed the false belief question if they both stated correctly their initial false belief about the contents of the box and responded correctly to the reality control question with what was actually in the box. All other responses were coded as incorrect.

Following an initial inspection of the data, children’s responses to the action explanation question were categorized as follows:

1. Correct Explanations:

Preaction: reference to a state that occurred prior to the children’s performing the action. These states included (a) false belief (e.g., “Because I thought it was crayons”), (b) desire (e.g., “Because I wanted a drink”), (c) a goal (e.g., “For drawing”), and (d) physical states (e.g., “Because I was thirsty,” “Because of crayons”).

1 Surprisingly, in only four trials across each of the four experiments reported in this article did children provide the response “Because you told me to” (referring to the fact that the experimenter had suggested that the child retrieve the paper [cup]). These four instances were coded as preaction (under the category of goal) explanations because they made reference to an event that had occurred prior to the child’s retrieval of the item.
2. Incorrect Explanations:
   Postaction: reference to a state that occurred after children had performed the action. Within this category, the only state that children referred to was the actual contents of the box (e.g., “Because there was sand in there”).
   Irrelevant: reference to an aspect of the situation that was not causally related to the child’s action (e.g., “I got the paper” or “Where’s the cup?”).
   No response: included trials in which the child provided no response at all or simply stated “I don’t know” or “Because.”
   Only explanations that were coded as preaction were considered to constitute a correct explanation for the child’s action (or, alternatively, a “pass” on the action explanation question). Included in this category were explanations that made reference to a false belief but also those that made reference to a desire, goal, or physical state of the world. In contrast, responses that were coded as postaction, irrelevant, or no response were considered incorrect (or, alternatively, a “fail” on the action explanation question). All of the children’s explanations were independently coded by Cristina M. Atance and by an undergraduate psychology student who was unaware of the purpose of the study. Interrater agreement was 96%, and Cohen’s kappa was .95. All disagreements were resolved through discussion.

Results

False Belief Question

Across both trials (crayon box and juice box), children passed the false belief question 50% of the time and failed 50% of the time. Children’s responses to the false belief question were not affected by trial order (Trial 1 vs. Trial 2) or by box type (crayon vs. juice).

Action Explanation Question

Analyses of children’s responses to the action explanation question were broken down as a function of their performance on the false belief question (i.e., whether they passed or failed this question). As shown in Table 1, on those trials in which children passed the false belief question ($n = 60$), 265% of their explanations were coded as preaction, 16% as postaction, 32% as irrelevant, and 20% as no response. Among the preaction explanations that children provided ($n = 39$), 15% ($n = 6$) appealed to a false belief, 31% ($n = 12$) to a desire, 39% ($n = 15$) to a goal, and 15% ($n = 6$) to a physical state (see Table 2 for examples of children’s explanations).

On those trials in which children failed the false belief question ($n = 63$), 19% of their responses to the action explanation question were coded as preaction, 16% as postaction, 32% as irrelevant, and 33% as no response. Among the preaction explanations that children provided ($n = 12$), 42% ($n = 5$) appealed to a false belief, 25% ($n = 3$) to a desire, 25% ($n = 3$) to a goal, and 8% ($n = 1$) to a physical state (see Table 3 for examples of children’s explanations).

Not surprisingly, there were significantly more preaction explanations provided on trials in which children passed the false belief question than on trials in which children failed the false belief question: Trial 1, $\chi^2(1, N = 62) = 6.22, p = .013$; Trial 2, $\chi^2(1, N = 61) = 22.89, p < .001$. Across the sample as a whole, the types of explanations that children provided did not differ as a function of trial order or box type.

Consistency Across Questions

We addressed several issues in these analyses. First, were children’s performances on the false belief and action explanation questions related? Second, was one question type (i.e., false belief or action explanation) significantly harder for children to pass? A Spearman rank order correlation revealed that children’s performances on the false belief and action explanation questions were indeed significantly related: $r(61) = .32, p = .012$, and $r(60) = .61, p < .01$, for Trials 1 and 2, respectively. Thus, children who passed the false belief question tended to pass the action explanation question. However, to determine whether children found one question more difficult than the other, we conducted McNemar chi-square tests. These analyses did not reveal a significant difference between children’s performances on these two questions. That is, children did not pass the false belief question any more often than they passed the action explanation question. If anything, children tended to find it slightly more difficult (though not significantly so) to correctly explain their action than to correctly retrieve their belief. In 35% of the trials, children passed the false belief question yet failed the action explanation question, whereas in 20% of the trials children failed the false belief question but passed the action explanation question (see Table 4).

Discussion

Our goal in Experiment 1 was to determine how 3-year-old children would explain their own false-belief-based action. Despite the fact that the children themselves carried out a goal-directed action premised on a false belief about the contents of a box, they often had difficulty explaining what motivated this action. Indeed, in 81% of instances in which children failed the false belief question, they were unable to correctly explain their action. This “explanatory” difficulty supports the results of previous research by Moses and Flavell (1990), Wimmer and Weichbold (1994), and Wimmer and Mayringer (1998) but represents the first empirical effort to extend these findings to the context of the child’s own

---

2 Because the experimental protocol did not initially entail asking children who passed the false belief question to explain their action, the total number of explanations provided in these cases was 60 rather than 63.
action. Finally, contrary to the results of Bartsch and Wellman (1989), children found it no easier to correctly explain their false-belief-based action than they did to retrieve their false belief. In fact, not all children who passed the false belief question went on to pass the action explanation question—an issue to which we return in the General Discussion. Moreover, as illustrated in Table 1, irrespective of whether children passed or failed the false belief question, it was rare that they explicitly referenced their false belief to explain their action.

One possibility is that children might have performed better if they had been asked to explain their action prior to being asked to state their belief. Although it was our intention to question children (who failed the false belief question) in a manner that would make the inconsistency between their stated belief (i.e., candles [sand]) and their prior action salient to them, it is possible that the results might have differed if the order of the questions had been reversed. That is, after having essentially denied holding a false belief, children may have been placed in an awkward position when asked to explain an action of theirs that was premised on this belief. Asking children the action explanation question prior to the false belief question would eliminate this possibility and might in fact serve to draw out more explicit references to false belief than were observed in Experiment 1. Thus, in Experiment 2, each child was given two false-belief-action tasks—in one of these, the false belief question was asked first (the false-belief-first task), as was the case in Experiment 1, whereas in the other task, the action explanation question was asked first (the action-explanation-first task). This allowed us to directly compare performance on each task.

### Experiment 2

#### Method

**Participants**

Participants were 28 three-year-olds (13 boys and 15 girls; age range = 3 years 1 month to 3 years 11 months; mean age = 3 years 7 months). One additional child was excluded from this final sample because of experimenter error. Most of the children were White and from middle-class backgrounds and were recruited by telephone calls from a university child studies participant pool.

**Materials**

Stimuli included a crayon box that contained candles, an orange juice box that contained sand, a sheet of 8½ × 11 in. white paper, and a blue plastic cup.

**Design and Procedure**

Children were tested individually in a laboratory playroom. Prior to the testing session, the piece of paper and the cup were placed alongside a wall adjacent to the table where the child and the experimenter would be seated during the session. The false-belief-first task was identical to the false-belief-action task in Experiment 1 and was administered in the same manner as in Experiment 1. The action-explanation-first task differed only in that the action explanation question (i.e., “Why did you go get the paper [cup]?”) was asked prior to the false belief question (i.e., “Before, when you first saw the box all closed up like this, what did you think was inside?”). For those children who correctly responded “crayons [juice],” the ensuing reality control question (i.e., “What is inside the box?”) was asked. Task order (action explanation first vs. false belief first) as well as box type (crayon vs. juice) were counterbalanced.

**Coding**

All of the children’s explanations were independently coded by Cristina M. Atance and by an undergraduate psychology student who was unaware of the purpose of the study. Interrater agreement was 95%, and Cohen’s kappa was .91. All disagreements were resolved through discussion.
Results

False Belief Question

When the results were collapsed across the two tasks (false belief first and action explanation first), children passed the false belief question 50% of the time and failed it 50% of the time. Children’s performance did not vary as a function of task (i.e., false belief first vs. action explanation first), task order, or box type.

Action Explanation Question

The main question that we sought to address in Experiment 2 was whether the order in which the action explanation and false belief questions were asked would influence the types of explanations that children provided for their prior action. That is, would asking children to explain their prior action prior to stating their belief increase the number of preaction (i.e., false belief, desire, goal, and physical state) explanations that they provided? A McNemar’s chi-square test revealed that children did not provide significantly more preaction explanations in the action-explanation-first task than in the false-belief-first task ($\chi^2 = 0.10, p = .75$). Overall, 46% of children’s explanations were coded as preaction in the action-explanation-first task, and 39% were coded as preaction in the false-belief-first task. Box type did not have an effect on the type of action explanations that children provided. Task order had the following effect: Children provided significantly more preaction explanations on the false-belief-first task when the explanation-first task preceded rather than followed it. However, to reiterate, children did not provide significantly more preaction explanations in one task than in the other.

Consistency Across Questions

The relation between children’s performance on the false belief question and their performance on the action explanation question was almost significant: $r(27) = .35, p = .067$, and $r(27) = .36, p = .061$, for the action-explanation-first task and the false-belief-first task, respectively. As in Experiment 1, McNemar chi-square tests did not reveal any significant difference between the number of correct responses that children provided to the false belief question and their performance on the action explanation question (see Table 5). In 35% of the trials, children passed the false belief question but failed the action explanation question, whereas in 29% of the trials children failed the false belief question but passed the action explanation question.

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Contingency Table for Children’s Responses to the Action Explanation and False Belief Questions (Collapsed Across Both Tasks) in Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>False belief question</td>
<td>Action explanation question</td>
</tr>
<tr>
<td>Fail</td>
<td>Pass</td>
</tr>
<tr>
<td>Fail</td>
<td>20</td>
</tr>
<tr>
<td>Pass</td>
<td>10</td>
</tr>
</tbody>
</table>

Discussion

The results of Experiment 2 demonstrate that children were no more likely to provide preaction explanations when they were asked to explain their prior action in question prior to being asked to state their belief than when they were asked to state their belief prior to being asked to explain their action. However, one remaining possibility for children’s difficulty concerns the general nature of answering why questions. More specifically, 3-year-olds may have difficulty formulating an explanation for any type of past action, even one that was not premised on a false belief. According to this logic, children should find it just as difficult to explain an action of theirs that was premised on a true belief as to explain one that was premised on a false belief. We explored this possibility in Experiment 3.

Experiment 3

Method

Participants

Participants were 26 three-year-olds (14 boys and 12 girls; age range = 3 years 0 months to 3 years 7 months; mean age = 3 years 2 months). Children were predominantly from White, middle-class families and were recruited by telephone calls from a university child studies participant pool.

Materials

Stimuli were a crayon box that contained crayons or candles (depending on task type), an orange juice box that contained orange juice or sand (depending on task type), a sheet of 8 1/2 × 11 in. white paper, and a blue plastic cup.

Design and Procedure

Children were tested individually in a laboratory playroom. Prior to the testing session, the piece of paper and the cup were placed alongside a wall adjacent to the table where the child and the experimenter would be seated during the session. Children were given one trial of each of the following tasks: (a) false belief action and (b) true belief action. The false-belief-action task was identical to that administered in Experiment 1. The true-belief-action task was as follows: Children were shown the crayon [juice] box and were asked to state what they thought was inside. Children were then asked to go get the paper [cup] so that they could draw with the crayons [drink some juice]. Children were then shown that the box contained crayons [juice]. Children were then asked a true belief question: “Before, when you first saw the box all closed up like this, what did you think was inside?” This was followed by the action explanation question: “Why did you go get the paper [cup]?” Thus, the procedure of the true-belief-action task was identical to that of the false-belief-action task, with the crucial difference that the content of the box was as children had expected (i.e., crayons or juice). The order of the two tasks along with box type (crayon vs. juice) was counterbalanced.

Coding

All of the children’s explanations were independently coded by Cristina M. Atance and by an undergraduate psychology student who was unaware of the purpose of the study. Interrater agreement was 98%, and Cohen’s kappa was .97. All disagreements were resolved through discussion.
Results

False Belief and True Belief Questions

In the false-belief-action task, 31% of the children passed the false belief question, whereas 69% failed this question. Neither task order nor box type had an effect on children’s responses to the false belief question. Not surprisingly, each child passed the true belief question in the true-belief-action task.

Action Explanation Question

Our main research question was whether children’s difficulty appealing to a false belief to explain their prior action was due to a general difficulty in explaining why questions or was due to the specific difficulty of explaining a prior action of theirs that was premised on a false belief. A McNemar’s chi-square test revealed that children provided significantly more preaction explanations in those instances in which they were asked to explain a prior action of theirs that was premised on a true belief than in instances in which they were asked to explain a prior action of theirs that was premised on a false belief. χ²(corrected, 1, N = 26) = 4.9, p = .021. Sixty-nine percent of children’s explanations were coded as preaction when they were asked to explain an action of theirs that was premised on a true belief, whereas only 38% of children’s explanations were coded as preaction when they were asked to explain an action of theirs that was premised on a false belief. Neither box type nor task order had an effect on children’s explanations.

Consistency Across Questions

We limited these analyses to the false-belief-action task. As in Experiment 2, the relation between children’s performance on the false belief and action explanation questions was almost significant, r(26) = .33, p = .10. For the false-belief-action task, children did not provide significantly more correct responses to the false belief question than to the action explanation question. Children passed the false belief question but failed the action explanation question in 37% of the trials, whereas children failed the false belief question but passed the action explanation question in 28% of the trials.

Discussion

Experiments 2 and 3 ruled out that children’s difficulty in explaining a false-belief-based action was due either to the order in which they were asked the false belief and action explanation questions or to a difficulty in answering why questions.

From a theoretical standpoint, children’s difficulty in explaining their action may reside in their inability to correctly identify their false belief as the motivator of their action (e.g., one’s belief that there are crayons in the box motivates one to get some paper to draw on). Thus, even though children were enthusiastic and willing to get the paper (as well as the cup) after the experimenter had suggested doing so to them, perhaps their explanatory competence would have been enhanced if they had played a more active role in this process. This might, in turn, have made the link between their belief and their ensuing action more salient. Thus, if prior to acting on their false belief, children were asked to formulate a plan of their own on the basis of this belief, it is possible that their explanatory competence might improve. We tested this prediction in Experiment 4.

In Experiment 4, the following planning + action task was devised. This task was similar to the false-belief-action task with the following exception: Prior to acting on their false belief, children were also asked to plan on the basis of their false belief. Thus, with the crayon box, for instance, children were asked to plan what they would draw and what color they would use for their drawing. We predicted that this “planning + action” context might lead children to provide a higher number of correct preaction explanations (including perhaps ones that made explicit reference to a false belief) than would be provided in the “action only” context.

In addition, we were interested in whether children’s performance on the false belief question would be enhanced by providing them with what can be characterized as increased evidence of their false belief. Thus, in addition to giving children the planning + action task and the false-belief-action task (hereafter referred to as the action task), children were given a standard unexpected contents task (e.g., Gopnik & Astington, 1988; hereafter referred to as the standard task), which provided a baseline measure of children’s ability to retrieve a false belief in a context that did not involve either a false-belief-based action or a false-belief-based plan. We predicted that children would retrieve their false belief most often in the planning + action task, and least often in the standard task, with performance on the action task falling somewhere in between.

Experiment 4

Method

Participants

Participants were 39 children (16 boys and 23 girls; age range 1 month to 3 years 11 months; mean age = 3 years 5 months). Seven additional children were excluded from the sample for either failing to state their initial belief about the contents of a box (n = 4) or not wanting to get the cup on a juice box trial (n = 3). All children were predominantly from White, middle-class families and were recruited from a university early childhood education center and a local day-care center.

Materials

The materials in this experiment included three boxes with unexpected contents—a crayon box that contained candles, a juice box that contained sand, and a Smiley box that contained string. Smarties are a candy very familiar to Canadian children and are similar to M&M’s. Materials also included three objects that the children were asked to retrieve: a piece of paper on which to draw with the crayons, a cup from which to drink the juice, and a bowl in which to put the Smarties.

Design and Procedure

Children were tested individually in a quiet area in their respective preschool settings. To reduce any carryover effects between the tasks, we tested children during two separate sessions that were no less than 2 days apart (e.g., Monday and Wednesday). Children were always given the planning + action task and the action task together in counterbalanced order during one session, and they were given the standard task during the other session. The order of these two sessions was counterbalanced such
that children were given either the planning + action task and the action task first and the standard task second, or the standard task first and the planning + action task and the action task second. The procedure of the tasks was as follows:

**Standard.** Children were shown one of the three boxes and were asked to state what they thought was inside. Children were then shown the unexpected contents of the box. The box was closed up and children were asked the false belief question: “Before, when you first saw the box all closed up like this, what did you think was inside?” Children who answered correctly were then asked the reality control question: “What is inside the box?”

**Action.** The procedure for the crayon and juice boxes was identical to the procedure in Experiment 1. For the Smarties box, children were asked to get a bowl in which to put the Smarties (i.e., “Look, there’s a bowl over there. Why don’t you get it to put the Smarties in?”). After discovering the true contents of the box, children were asked the following action explanation question: “Why did you go get the bowl (then)?”

**Planning + action.** The procedure for this task was identical to the procedure for the action task with the following exception: Prior to being asked to perform an action, children were asked to plan what they would do with the expected contents of the box. For the crayon box, children were asked, “What are you going to draw?” and “What color are you going to use to draw X?” For the juice box, children were asked, “Can you show me how much you’re going to drink?” (children pointed to the level on the box corresponding to how much they would drink) and “How many sips are you going to take?” Finally, for the Smarties box, children were asked, “How many are you going to eat?” and “What color are you going to eat first?” The administration of the test questions in this task was identical to their administration in the action task.

**Coding**

The verbal coding procedure for Experiment 4 was identical to that of Experiment 1. Interrater agreement for children’s explanations was 100%.

**Results**

**False Belief Question**

Children’s responses to the false belief question did not differ significantly as a function of task order. Overall, children passed the false belief question 59% of the time in the planning + action task, 54% of the time in the action task, and 46% of the time in the standard task. Using a one-tailed significance level (because specific predictions about children’s performance had been made), a McNemar chi-square test revealed a significant difference between children’s performance on the planning + action task and their performance on the standard task, \( \chi^2(1, N = 39) = 3.2, p = .032 \) (see Figure 1). In other words, as predicted, children passed the false belief question significantly more often in the planning + action task than in the standard task. No other significant task differences emerged.

Of the 39 children in the sample, 15 passed the false belief question across all three tasks, whereas 12 failed the false belief question across all three tasks. The pattern for the remaining 12 children is shown in Table 6. As shown in this table, it was never the case that children answered the false belief question correctly in the standard task without also doing so in the planning + action task.

**Action Explanation Question**

As expected, children provided significantly more preaction explanations (i.e., belief, desire, goal, or physical state) on those tasks in which they answered the false belief question correctly than on those in which they did not. This was the case for both the planning + action task, \( \chi^2(1, N = 39) = 8.20, p = .004 \), and the action task, \( \chi^2(1, N = 39) = 10.09, p = .001 \). Contrary to what we had predicted, children did not provide a greater number of preaction explanations in the planning + action task than in the action task.

**Consistency Across Questions**

Children’s performance on the false belief and action explanation questions was significantly related. For both the action task and the planning + action task, the Spearman rank order correlation was significant, \( r(38) = .46, p = .003 \). Again, no significant differences were obtained between children’s performance on the action explanation question and their performance on the false belief question. However, as in each of the previous experiments, children tended to have slightly more difficulty correctly explaining their action than correctly stating their initial belief (see Table 7). Children passed the false belief question but failed the action explanation question in 35.5% of the trials, whereas children failed the false belief question but passed the action explanation question in 15% of the trials.

**Pattern of Findings Across Experiments 1 to 4**

We analyzed children’s performance on the action explanation question across each of the four reported experiments to determine whether the number of explanation types provided differed. We collapsed across explanation types to look specifically at the following three: (a) false belief explanations; (b) all other preaction explanations (i.e., desire, goal, and physical state); and (c) postaction, irrelevant, and no response explanations. These analyses revealed that there were no differences in the distribution of these explanation types across experiments. For instance, the children in Experiment 4 did not provide any more explanations that made explicit reference to a false belief than did the children in
Experiment 1 (see Table 8 for a summary of the results across experiments).

Discussion

As predicted, children passed the false belief question significantly more often in the planning + action task than in the standard task. Children’s performance on the false belief question was not significantly better in the action task than in the standard task. Thus, it appears that simply acting on a false belief does not enhance retrieval of a false belief, no research has addressed how they explain someone else’s false-belief-based action than they did stating this same person’s false belief, but it is consistent with the findings by Moses and Flavell (1990), Wimmer and Weichbold (1994), and Wimmer and Mayringer (1998), who did not find explanation to be easier for children.

Table 6

<table>
<thead>
<tr>
<th>Task(s) in which false belief question was answered correctly</th>
<th>No. of children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard only</td>
<td>0</td>
</tr>
<tr>
<td>Action only</td>
<td>4</td>
</tr>
<tr>
<td>Planning + action only</td>
<td>3</td>
</tr>
<tr>
<td>Standard and action</td>
<td>0</td>
</tr>
<tr>
<td>Standard and planning + action</td>
<td>3</td>
</tr>
<tr>
<td>Action and planning + action</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7

Contingency Table for Children’s Responses to the Action Explanation and False Belief Questions (Collapsed Across Action and Planning + Action Tasks) in Experiment 4

<table>
<thead>
<tr>
<th>Action explanation question</th>
<th>Fail</th>
<th>Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>Pass</td>
<td>16</td>
<td>28</td>
</tr>
</tbody>
</table>

Children’s ability to explain their false-belief-based action was not enhanced by a context that required them to plan and act on the basis of their belief (i.e., the planning + action task), compared with a context in which they only performed an action (i.e., the action task). Interestingly, however, children’s ability to retrieve a false belief was enhanced by this planning + action context, compared with a context in which they neither planned nor acted (i.e., the standard task). This last finding fits nicely within the framework of the theory theory. According to the theory theory, young children may only appeal to a false belief in those instances in which they have been exposed to extremely salient counterevidence (Astington & Gopnik, 1991; Gopnik, 1996a, 1996b; Gopnik & Meltzoff, 1997; Gopnik et al., 1994; Gopnik & Wellman, 1992). In the case of the planning + action task specifically, children likely had a recollection of having planned a course of action (e.g., “I’m gonna draw a car, and I’m going to use blue”) as well as of acting to fulfill this plan (e.g., go get the paper). Such an active involvement on behalf of the children may have led to an elaboration of their belief, thus rendering it more salient to them. This, in turn, may have made it more difficult for children to state that their initial belief was that the box contained candies. Such an answer is incongruous with evidence that points to the contrary (i.e., the child’s plan and action). Arguably, then, situating the child’s belief in a web of interconnected plans and actions may help him or her to retrace a prior state of mind and, more specifically in this case, a false belief. This interpretation is consistent with work by Mitchell and Lacoheé (1991), who found that if children were provided with a physical instantiation of their belief (in this case, a picture of Smarties to put into a mailbox), they could more easily retrieve this belief. Our findings extend this idea as they suggest that behavioral evidence (i.e., acting and planning) also seems to aid children in the retrieval process. What our findings cannot allow us to determine is whether such an effect would also carry over to a situation in which someone other than the child generated a plan about what to do with the contents of the box. That is, is it necessary for the child to play an active role in the generation of a plan, or is it sufficient for someone else to generate it?

At this point, we shift our focus back to the realm of action explanation. The main issue that we address is the difficulty that children experienced with action explanation. Recall that this difficulty applied even to those children who passed the false belief question. Indeed, in Experiment 1, 35% of the trials in which children passed the false belief question were not accompanied by correct responses to the action explanation question (i.e., making reference to a prior belief, desire, goal, or physical state). This figure was essentially replicated in Experiments 2, 3, and 4 (36%, 40%, and 35%, respectively).
37%, and 35.5%, respectively). Why might this have been the case?

It is important to first rule out a purely methodological reason for this finding. Specifically, one could argue that the children who fit this profile were those who answered the false belief question correctly solely by chance and thus should not be characterized as having an understanding of false belief. These children’s difficulty with false belief simply became apparent when they were asked to reason about an action of theirs that was premised on this belief. However, this interpretation is unlikely given that children were not given forced-choice response options to the false belief question but rather had to retrieve the correct response on their own. Moreover, some have argued that for the child who does not have an understanding of false belief, making a reality-based error (e.g., choosing the reality-based response “candles”) is the most likely option (Perner, 1995; Saltmarsh, Mitchell, & Robinson, 1995; Wellman et al, 2001). Thus, we believe that it is unlikely that the majority of the children who answered the false belief question correctly did so strictly by chance. Instead, we would argue that the children who fit this profile were unable to make sense of their prior action because they did not possess a fully formed understanding of false belief. Although they may have been able to understand that they had entertained a false belief, they may have had difficulty understanding how this false belief influenced their prior action. In what follows, we attempt to clarify why this might have been so.

Lee and Homer (1999) listed three fundamental assumptions that we hold about the mind: (a) The mind exists; (b) the mind consists of various mental states, including intention, knowledge, and belief; and (c) there exist causal links between the mind, external environments, and actions (the causality assumption). This last assumption is particularly relevant to the current discussion as it implies that it is possible to assess children’s representation of another person’s (or their own) behavior. Indeed, this was the case when the children in our experiments were asked the false belief question. In this context, children needed only to retrieve their earlier belief without also having to reason about how this belief influenced their behavior. However, to explain their prior action, it was imperative that children represent the causal link between an earlier context (e.g., drawing) and their subsequent action that was premised on this context. Thus, it is possible that not only false-belief-related explanations, but also desire-, goal-, and physical-state-related explanations may have required some understanding of the causal nature of false belief. That is, one could argue that implicit in the explanations “I wanted to draw,” “To color,” and “Because of crayons” is an understanding of the link between a prior belief and a subsequent action. Determining whether such explanations are indeed “implicit” references to false belief is beyond the scope of this article. However, suffice it to say that explanations of this type would appear to be at least a minimal requirement for granting a child a complete understanding of false belief and, more specifically, an understanding of the causal properties of false belief. Accordingly, it does not seem valid to attribute an understanding of false belief to a child who is unable to provide such explanations even though he or she might be able to retrieve a false belief.

Finally, some may argue that only explicit references to a false belief are clear evidence that the child understands that the ultimate cause of her prior action was her false belief. If so, then our data would force us to conclude that very few 3-year-olds indeed have this understanding, given that children—whether they passed or failed the false belief question—did not often make explicit references to a false belief (e.g., “Because I thought there were crayons in the box”) in their explanations (see Table 8). Consequently, we would conclude that explaining a prior action of one’s own by making explicit reference to a false belief is actually much harder than retrieving a false belief.

Thus far we have been alluding to the fact that children’s causal reasoning is related to their ability to explain a false-belief-based action. Although a number of studies indicate that in the physical

---

Table 8

<table>
<thead>
<tr>
<th>Coding category</th>
<th>Experiment 1 (n = 123)</th>
<th>Experiment 2 (n = 56)</th>
<th>Experiment 3 (n = 26)</th>
<th>Experiment 4 (n = 78)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass FBQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False belief</td>
<td>10 (6)</td>
<td>25 (7)</td>
<td>38 (3)</td>
<td>24.5 (11)</td>
</tr>
<tr>
<td>Desire</td>
<td>20 (12)</td>
<td>4 (1)</td>
<td>0 (0)</td>
<td>9 (4)</td>
</tr>
<tr>
<td>Goal</td>
<td>25 (15)</td>
<td>28 (8)</td>
<td>25 (2)</td>
<td>27 (12)</td>
</tr>
<tr>
<td>Physical</td>
<td>10 (6)</td>
<td>7 (2)</td>
<td>0 (0)</td>
<td>4 (2)</td>
</tr>
<tr>
<td>Postaction</td>
<td>5 (3)</td>
<td>7 (2)</td>
<td>0 (0)</td>
<td>4 (2)</td>
</tr>
<tr>
<td>Irrelevant</td>
<td>10 (6)</td>
<td>4 (1)</td>
<td>12 (1)</td>
<td>7 (3)</td>
</tr>
<tr>
<td>No response</td>
<td>20 (12)</td>
<td>25 (7)</td>
<td>25 (2)</td>
<td>24.5 (11)</td>
</tr>
<tr>
<td>Fail FBQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False belief</td>
<td>8 (5)</td>
<td>4 (1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Desire</td>
<td>5 (3)</td>
<td>7 (2)</td>
<td>11 (2)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Goal</td>
<td>5 (3)</td>
<td>14 (4)</td>
<td>11 (2)</td>
<td>12 (4)</td>
</tr>
<tr>
<td>Physical</td>
<td>2 (1)</td>
<td>4 (1)</td>
<td>6 (1)</td>
<td>3 (1)</td>
</tr>
<tr>
<td>Postaction</td>
<td>15 (10)</td>
<td>21 (6)</td>
<td>22 (4)</td>
<td>31 (10)</td>
</tr>
<tr>
<td>Irrelevant</td>
<td>32 (20)</td>
<td>14 (4)</td>
<td>17 (3)</td>
<td>21 (7)</td>
</tr>
<tr>
<td>No response</td>
<td>33 (21)</td>
<td>36 (10)</td>
<td>33 (6)</td>
<td>33 (11)</td>
</tr>
</tbody>
</table>

Note. Values enclosed in parentheses represent individual frequencies. FBQ = false belief question.
realm children are sensitive to the fact that a cause must precede its effect (e.g., Bullock & Gelman, 1979; Gopnik, Sobel, Schulz, & Glymour, 2001), a series of studies conducted by Povinelli and his colleagues (Povinelli, Landau, & Perilloux, 1996; Povinelli, Landry, Theall, Clark, & Castille, 1999) reveals a rather striking limitation in one specific aspect of preschool children’s understanding of causality. In these studies, 2-, 3-, and 4-year-old children were videotaped while playing a game with an experimenter. During the videotaping, unbeknownst to the child, the experimenter covertly placed a sticker on the child’s head. Several minutes later, the child was shown a video playback of the previous events, including the segment in which the experimenter placed the sticker on the child’s head. The researchers were interested in determining whether children would reach up to remove the sticker. Results indicated that none of the 2-year-olds and only about 25% of the 3-year-olds did so. Only by 4 years of age did a substantial number of children (75%) do so (Povinelli et al., 1996).

Povinelli (2001) argued that the findings from this delayed self-recognition task suggest that 2- and 3-year-old children do not understand how recent past events that the self has experienced are causally connected to the self’s current experiences. That is, in the delayed self-recognition task, these children seemed unable to understand how a past event (stickers being put on their heads) bore any relation to their current selves. Similarly, one could argue that the 3-year-old children in our experiments had difficulty conceptualizing how a prior state of the world, which was no longer true, could have caused their action. Indeed, Barresi (2001) has recently argued that children’s performance on false belief tasks is related to an understanding of the concept that the self extends through time and, more specifically, the concept that mental states of the self can change over time. Barresi (2001) stated that “to understand mental states as representational rather than as presentational states of the world, or of the organism, it may be necessary to conceive of the individual mind as existing outside of a particular time” (p. 157). Thus, part of acquiring an understanding of false belief may entail the ability to acknowledge how successive states of the self (physical or mental) are causally connected through time. In light of both Povinelli’s (2001) and Barresi’s (2001) claims, it is plausible that the action explanation question provided a more stringent test of children’s understanding of the causal structure of false belief than did the false belief question on its own. Alternatively, recent research by Povinelli and Dunphy-Lelii (2001) has shown that there is a marked increase between the ages of 3 and 5 years in children’s drive to seek explanations when an unexpected event occurs. This weaker drive to explain events (at least those that are unexpected) on the part of younger preschoolers may also have played a role in the explanatory difficulties observed in our experiments.

The current research has shown that explaining their own false-belief-based actions is difficult for 3-year-olds. Despite the fact that children acted on a belief, some were unable to identify this belief as the cause of their action. Moreover, explaining their false-belief-based action was no easier for children than retrieving their false belief. These experiments are the first to explore this important facet of young children’s reasoning about false belief, and thus we draw some important implications from them. First, our results show that children who are unable to retrieve a false belief (when asked a standard false belief question) do not suddenly appeal to this belief when asked to explain an action of their own that was premised on it. This is despite the fact that making sense of one’s action rests on the crucial understanding that it was a prior state of the world, not a current state, that motivated the action. Even when children were asked to explain their action prior to being asked to state their belief, as was the case in Experiment 2, they still experienced difficulty. Thus, it is not possible that children’s difficulty was a methodological artifact of being asked to state their belief prior to being asked to explain their action. And the results of Experiment 3 further show that the children’s difficulty could not be characterized as a general difficulty in answering why questions or in explaining any type of prior action (e.g., one premised on a true belief).

The findings from these studies underscore the importance of examining various aspects of young children’s understanding of false belief. For instance, it is important to be aware that although a child may correctly answer a false belief question, this same child may be unable to correctly reason about how this false belief influenced his or her behavior. And, similarly, there are instances in which a child may not be able to retrieve his or her false belief but may be able to explicitly appeal to this false belief to explain his or her action (though these instances are much rarer). This distinction highlights the need to situate children’s understanding of false belief within a broader conceptual framework. Indeed, characterizing a child as having an understanding of false belief may not be an accurate assessment if this child can pass a false belief question but cannot go on to reason about the behavioral implications of this belief. In contrast, if another child can pass a false belief question as well as correctly explain an action of his or hers that was premised on this belief, then one can characterize that child as having a more complete understanding of false belief than the former child. Finally, these findings highlight the need to frame false belief understanding in the context of such abilities as causal reasoning and, more specifically, an understanding of how the self’s actions are causally connected through time.

References


