The Curse of Knowledge in Reasoning About False Beliefs

Susan A. J. Birch and Paul Bloom
Yale University

Young children have problems reasoning about false beliefs. We propose that these problems arise, at least in part, because they possess an exaggerated version of the same bias that exists in adults - the 'curse of knowledge', which is a tendency to over-attribute one's own knowledge to someone else. If so, then even adults should be biased by their own knowledge when reasoning about false beliefs. This prediction was supported in two experiments using a modified false belief task. We also found that adults' own knowledge only led to biased predictions in the presence of a potential justification for their response (Experiment 1) and that this bias cannot be explained by salience alone (Experiment 2). These data support the more general claim that there is a fundamental continuity in children's and adult's ability to reason about the mind.

Over the last two decades, a wealth of research has explored the development of false belief reasoning in young children (see Wellman, Cross, & Watson, 2001 for a review). As Dennett (1978) pointed out, it is possible to correctly predict the action of another individual without any understanding of mental states, by simply observing the true state of the world — as long as the other individual's belief of the world is consistent with reality. A more valid test of whether someone appreciates mental states involves predicting actions when the other individual's mental state is inconsistent with reality — thus, the interest in children's ability to reason about false beliefs.

Most research on false belief reasoning has utilized some variant of the 'displacement task' (e.g. Wimmer & Perner, 1983; Baron-Cohen, Leslie, & Frith, 1985). For example, participants are told a story about Sally who puts her candy in a basket and leaves. In her absence, another character moves the candy to a box. When Sally returns, where will she look for her candy? The right answer, the basket, requires attributing a false belief to Sally. Another method is the 'Smartsies task', developed by Perner, Leekam, and Wimmer (1987). Children are shown a closed Smartsies container and asked what is inside. When they give the answer, “Smartsies”, the experimenter opens the container to reveal that there are actually pencils inside. The box is then closed and the child is asked what someone else, who was absent when the pencils were revealed, will think is inside. The child is also asked what he or she originally said was inside. Again, getting these questions right requires attributing a false belief, either to oneself or to someone else.

Four-year-olds tend to do well at these tasks, but younger children tend to fail (see Wellman et. al.’s, 2001 meta-analysis of 178 studies). The younger children do not answer randomly; instead, they tend to answer in accord with their own knowledge, saying that Sally will think the candy is in the box, that the other person will think that there are pencils in the container, and that they themselves had originally said the container held pencils.

The source of this difficulty is a matter of considerable debate. Some see the capacity to reason about the minds of others as the product of a uniquely human capacity to develop intuitive theories (e.g., Gopnik, 1996; Gopnik & Meltzoff, 1997) or as a product of cultural learning (e.g., Lillard, 1998). From this perspective, the child's problem may well be a bona fide conceptual one, revealing a genuine difference in how they understand mental life. A different possibility is that the capacity for belief attribution is largely unlearned, and present in at least a rudimentary form at a young age (Fodor, 1992; Leslie, 1987). From this perspective, young children's problem in reasoning about false beliefs might be due to more general factors such as memory load, assumptions of rationality, conflicting representations, processing limitations, among others, and thus not necessarily indicative of a different concept of belief (Germain & Leslie, 2000; Koos, Gergely, Csibra, & Biro, 1997; Moses, 1993; Roth & Leslie, 1998; Zaitchik, 1990; for a discussion, see Bloom & German, 2000).

The proposal we explore here is most consistent with the view that children’s early difficulties do not stem from a lack of understanding of beliefs per se. But we offer a unique explanation for children’s errors. We propose that children’s difficulties might be accounted for, at least in part, by a greater susceptibility to a bias found in adults. Economists and social psychologists have pointed out that adults suffer from the curse of knowledge (Camerer, Weber, & Lowenstein, 1989; sometimes referred to as ‘hindsight bias’ e.g. Fischhoff, 1975), a tendency to be biased by one’s own knowledge when attempting to take a more naïve perspective. Once we know the solution to a problem, we

This research was supported by a Natural Sciences and Engineering Research Council of Canada Fellowship awarded to the first author.

Correspondence can be addressed by mail to: Susan Birch, Department of Psychology, Yale University, PO Box 208205, New Haven, CT, 06520; by e-mail to: susan.birch@yale.edu, or by phone (203) 436-1545.
tend to overestimate how easy it is for someone else to solve (Kelley & Jacoby, 1996). Similarly, if we know a company's earnings (Camerer, et. al., 1989), the meaning of an idiom (Keysar & Bly, 1995), or whether a statement is sarcastic (Keysar, 1994), our assessments of the judgments of a naïve person are biased in the direction of what we know (see also Fischhoff, 1975; Keysar, Ginzel, & Bazerman, 1995; Newton, 1990).

Not surprisingly, children are also susceptible to the curse of knowledge (Bernstein, Atance, Loftus, & Meltzoff, in press; Birch & Bloom, 2003). In Birch & Bloom (2003), 3- to 5-year-old children were presented with novel toys and were told that each contained “something special” inside. One toy was presented as familiar to the experimenter’s puppet friend Percy (i.e., he said: “I’ve played with that one before”). The other was presented as unfamiliar to Percy (i.e., he said: “I’ve never ever seen that one before”). The key manipulation was whether the children themselves knew what was inside the toys. For half of the trials, the children were ‘cursed’ by showing them what was inside the toys prior to Percy’s appearance. For the other half of the trials, the children were not shown what was inside the toys. The results revealed a significant curse of knowledge bias: When the children themselves knew what was inside they were more likely to over-attribute knowledge to Percy. The opposite is not true, however: when the children were ignorant about what was inside they were not more likely to under-attribute knowledge to Percy. There is no corresponding ‘curse of ignorance’, suggesting that the results cannot be accounted for by a more general egocentric bias (see Birch & Bloom, 2003 for discussion). Particularly noteworthy is that the magnitude of the bias significantly declined from age 3 to age 5 — the same age children begin succeeding at false belief tasks.

Importantly, traditional false beliefs tasks are ‘cursed’ — participants know the outcome and are asked to judge how a more naïve other (or themselves previously) would think or act. Indeed, there are many parallels between curse-of-knowledge studies with adults and the false belief tasks given to children. First, consider an experiment by Fischhoff (1975) who provided adult participants with descriptions of events and told them that these descriptions were also presented to other students. Participants who were told the outcome of the events estimated that naïve students would predict higher probabilities to that outcome than participants who were ignorant of the outcome. In other words, like children in the displacement task who responded as if naïve Sally shared their outcome-knowledge, adults in Fischhoff’s (1975) task who were told the event outcome also responded as if naïve others shared their knowledge.

Second, consider Fischhoff and Beyth (1975)’s experiment in which participants were asked on the eve of former President Nixon’s trips to China and the USSR to estimate the probability of various possible outcomes of the visits. After the trip’s completion, the same participants were asked to recall their original predictions. Like children in the Smarties task who tended to report their new knowledge of what was in the box instead of their original predictions, participants’ recollections were biased in the direction of their knowledge of what had actually taken place.

To explore these parallels in a more direct fashion, we gave adults a displacement task that differed from standard tasks in three important ways. First, we used a more sensitive measure than the categorical response that is typically obtained. Participants were asked to report the probability that the protagonist would look in each of the containers when she returned. Second, instead of two containers, we used four, to enable us to manipulate the participants’ knowledge of the outcome. Either participants were told the violin was moved to a specific container; or told that the violin was moved to another container but were not told which one (i.e. the ‘Ignorant’ condition). We predicted that knowledge of the outcome in a false belief task like that used with children would compromise even adults’ ability to reason about false beliefs. Third, the containers were rearranged following displacement to allow us to manipulate the plausibility that the protagonist would look in each of the containers. In a ‘Knowledge – Plausible’ condition, participants were told that the violin was moved to a container that, following rearrangement, was in the same location as the violin was originally—thus, it would be reasonable to assume that the character might look there by mistake. In a ‘Knowledge – Implausible’ condition, participants were told that the violin was moved to a container that following rearrangement was in a different location than where the violin was originally. A ‘True Belief’ condition in which the violin was not moved was also included. We predicted that knowledge of the outcome might only be a curse if the participants had some reason to justify their responses.

**Experiment 1**

**Method**

**Participants.** Two hundred ten students enrolled in an Introductory Psychology course at Yale University in New Haven, CT participated. Ninety-three were male, 112 were female, and 5 opted not to specify their gender. An additional 17 participants (7.5%) were excluded because their percentage totals did not equal 100%.

**Materials and Procedure.** Participants were randomly assigned to one of four conditions: True Belief, Ignorant, Knowledge – Plausible, and Knowledge – Implausible. All participants received the same stimuli—a color version of Figure 1 (without the color word labels) depicting a girl holding a violin by a sofa and four containers. Each container was a different color: blue, purple, red, and green. Beneath the first picture was an image of a different girl holding a violin and the same four containers rearranged: red, green, purple, blue.

Participants in all four conditions read, This is Vicki. She finishes playing her violin and puts it in the blue container. Then she goes outside to play. While Vicki is outside playing her sister, Denise…” At this point, the conditions differed:

- **True Belief:** “comes into the room.”
- **Ignorant:** “moves the violin to another container.”
- **Knowledge – Plausible:** “moves the violin to the red container.”
- **Knowledge – Implausible:** “moves the violin to the purple container.”
All participants then read, “Then, Denise rearranges the containers in the room until it looks like the picture below.” This was followed by, “When Vicki returns she wants to play her violin. What are the chances Vicki will first look for her violin in each of the above containers? Write your answers in percentages in the spaces provided under each container.”

![Image](34x378 to 303x648)

Figure 1 The Knowledge – Plausible Version of the Stimuli

Results and Discussion

In the True Belief condition participants gave a mean probability rating of 68.45% (SD = 25.57%) to the blue container where Vicki placed the violin and a mean probability rating of 28.44% (SD = 23.59%) to the red container that following rearrangement occupied the location that the blue container previously occupied. That is, most of the time they thought Vicki would look in the original container, but often they thought she would look in a different container that was in the same location as the violin was originally. They gave the other two containers, not the container that held the violin, and not the same location either, a combined probability rating of just over 3%. See Table 1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Container</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blue</td>
</tr>
<tr>
<td>True Belief</td>
<td>68%</td>
</tr>
<tr>
<td>Ignorant</td>
<td>71%</td>
</tr>
<tr>
<td>Knowledge – Plausible</td>
<td>59%</td>
</tr>
<tr>
<td>Knowledge – Implausible</td>
<td>73%</td>
</tr>
</tbody>
</table>

Table 1 Mean Probability Judgments that Vicki Will Look in Each of the Containers

In the Ignorant condition, probability ratings that Vicki would look for her violin in the blue container where she left it (M = 71.39%, SD = 25.74%), did not differ from that of the True Belief condition, t (104) = -0.59, p = .56, NS (two-tailed). That is, participants who knew that the violin was moved but did not know where it was moved, predicted that it was as likely that Vicki would act based on a false belief, as participants who knew the violin had never been moved.1

In contrast, consider the Knowledge – Plausible condition in which participants were told that the violin was moved to the red container, a container in a location that Vicki might plausibly look. Here, participants reported significantly higher probabilities to the red container, than participants in the Ignorant condition, t (105) = -2.42, p < .05 (one-tailed). In addition, participants in the Knowledge – Implausible condition reported significantly lower probabilities that Vicki would look in the blue container (M = 59.38%, SD = 26.99%) than participants in the Ignorant condition (M = 71.39%, SD = 25.74%), t (105) = 2.35, p < .05 (one-tailed). Also, participants’ knowledge that the violin was in the purple container did not lead them to decrease their probability judgments that Vicki would possess, or act on, a false belief. In other words, participants reported similar probabilities for the blue container in both the Knowledge – Implausible (M = 72.56%, SD = 28.58%) and Ignorant conditions (M = 71.39%, SD = 25.74%), t (97) = -0.21, p = .42, NS (one-tailed).

To sum up, our findings demonstrate that a person’s own knowledge can interfere with their ability to reason about another’s false belief. However, in the false belief task we explored, the participants’ knowledge only led to biased responses when there was a seemingly justifiable reason to support such a response.

Experiment 2

What role does salience play in the previous findings? Does the participant have to believe that the violin is in a specific container, or is merely drawing attention to that container enough to increase participant’s probability judgments for that container? One could argue that simply by mentioning the red container, and thus drawing attention to it, participants would be more likely to provide higher probabilities to that container. Alternatively, it is not the

---

1 There was a significant difference between the ratings for the purple container in the True Belief (M = 1.00%, SD = 2.32%) and Ignorant conditions (M = 2.46%, SD = 4.56%), t (104) = -2.103, p < .05, two-tailed. We did not predict this finding, nor do we have a good explanation for it.
salience of the red container per se that matters but the knowledge that the red container is where the object is hidden. We predicted that merely mentioning the red container would not be sufficient to significantly increase participants’ assessments of the likelihood Vicki would look there, nor would it significantly bias their false belief reasoning.

Method

Participants. One hundred ten students enrolled in an Introductory Psychology course at Yale University in New Haven, CT participated. Fifty were male, 59 were female, and 1 person opted not to specify his or her gender. An additional 12 participants (9.8%) were excluded because their percentage totals did not equal 100%.

Materials and Procedure. Participants were given the same pictorial representation of the displacement task utilized in Experiment 1. Participants were randomly assigned to one of three conditions: Knowledge (Plausible), Ignorant, and Salience. The task was the exact same as in Experiment 1 except after reading, “This is Vicki. She finishes playing her violin and puts it in the blue container. Then she goes outside to play. While Vicki is outside playing her sister, Denise…” the conditions differed as follows:

Knowledge (Plausible): “moves the violin to another container—the red one.”

Ignorant: “moves the violin to another container.”

Salience: “moves the violin to another container—not the red one.”

Results and Discussion

Replicating our main finding from Experiment 1, we found that participants in the Knowledge condition reported significantly higher probabilities that Vicki would look in the container that they knew held the violin (M = 33.51%, SD = 24.09%), compared to participants in the Ignorant condition who did not have specific knowledge about the violin’s whereabouts (M = 19.49%, SD = 21.88%), t(68) = -2.52, p < .05 (one-tailed). These ‘cursed’ participants also reported significantly lower probabilities that Vicki would look in the container she should falsely believe held her violin (M = 60.23%, SD = 26.34%), compared to participants in the Ignorant condition (M = 74.06%, SD = 27.72%), t(68) = 2.13, p < .05 (one-tailed). Thus, reproducing our finding that a person’s own knowledge can interfere with their ability to appreciate someone else’s false belief, even on a seemingly simple task like that used with children. See Table 2.

The main question of interest in Experiment 2 concerns the salience condition. Participants’ probability judgments that Vicki would look in the blue container in the Salience condition were not significantly different from either the Knowledge (M = 60.23%, SD = 26.34%), t (77) = -1.30, p = .10, NS (one-tailed) or Ignorant conditions (M = 74.06%, SD = 27.72%, respectively) t (69) = -30 NS (two-tailed). However, participants in the Salience condition reported significantly lower probabilities that Vicki would look in the red container (M = 24.93%, SD = 21.57%), than participants in the Knowledge condition (M = 33.51%, SD = 24.09%), t (77) = 1.67, p < .05 (one-tailed). Furthermore, participants in the Salience condition did not report significantly higher probabilities that Vicki would look in the red container (M = 24.93%, SD = 21.57%), compared to participants in the Ignorant condition (M = 19.49%, SD = 21.88%), t (69) = -1.05, p = .30, NS (two-tailed). These data suggest that salience cannot account for the bias in adults’ false belief reasoning in its entirety.

General Discussion

Our findings reveal that a person’s own knowledge can contaminate his or her ability to reason about false beliefs. In Experiment 1, we demonstrated that a person’s own knowledge can bias one’s ability to reason about another’s false belief but that this bias is affected by plausibility. In the condition in which the ‘cursed’ container was in a plausible location for Vicki to look, participants’ assessments of the likelihood that Vicki would hold a false belief were significantly reduced. However, in the condition in which the ‘cursed’ container was in an improbable location, participants’ judgments of the probability that Vicki would hold a false belief did not decrease.

In Experiment 2, we replicated the finding that adults’ ability to attribute false beliefs can be biased by their own knowledge. Adults who knew exactly where Vicki’s violin was moved reported significantly lower probabilities that Vicki would hold a false belief than adults who did not possess such knowledge. Experiment 2 also demonstrated that drawing attention to the container is not sufficient to increase participants’ probability judgments to that container or to interfere with their false belief reasoning. Salience may contribute to the biasing effects of knowledge but cannot account for the bias in and of itself.

The findings from Experiment 1 have implications for the study of adult social cognition. Due to the nature of the tasks that have typically been used in the adult literature, participants have only been ‘cursed’ with plausible or reasonable knowledge (e.g. the outcomes of Nixon’s visits, whether or not statements are sarcastic, solutions to problems, etc.). It might be that the results would be substantively different if the participants were told about event outcomes that were less plausible – perhaps the curse of knowledge would be diminished or would disappear. Our findings suggest that knowledge is a more potent curse when it is in conjunction with a potential rationale for inflating one’s estimates of what others know.

More generally, these findings have implications for research on children’s and adults’ ability to reason about the mental states of others. Anytime someone is more
knowledgeable than the person whose perspective they are trying to take, they are subject to the curse of knowledge. In this way, the curse of knowledge is importantly different from ‘egocentric’ accounts. That is, if someone is more ignorant than the person whose perspective they are trying to take, they are not subject to the symmetric ‘curse of ignorance’ counterpart (Birch & Bloom, 2003). More specifically, however, these findings show how the curse of knowledge can hinder one’s ability to reason about false beliefs and demonstrate how the contaminating effects of knowledge contribute an added source of difficulty to false belief tasks.

In the course of exploring other hypotheses, other researchers have somewhat reduced the ‘curse of knowledge’ in false belief tasks (Wimmer & Perner, 1983; Koos et al., 1997). For example, the experimenter ate the candy, or the antagonist in the story baked it into a cake. Consistent with our findings, children did significantly better under these conditions than in the standard task. However, the youngest children’s performance is still not significantly above chance (Wellman et al., 2001).

There are two possible explanations for why young children’s performance might improve under such conditions but would still not exceed chance. It may be that even if the curse of knowledge were removed young children would still experience difficulty with false belief tasks because of other problems posed by the task (see Bloom & German, 2001; Fodor, 1992; Leslie, 1994; Zaitchik, 1990) or due to a genuine inability to reason about false beliefs (see Gopnik, 1993; Perner, 1991; Wellman et al., 2001).

Alternatively, it may be that in these studies the ‘curse of knowledge’ is not adequately removed. Knowledge, after all, is graded. Knowing that the candy has been moved to one of three containers, but not knowing which one, is less specific than knowing that it has been eaten, but more specific than knowing it was moved to one of seven possible containers, for example. Simply knowing that it has been moved at all means one is not completely ignorant. We propose that the more specific the knowledge, the more potent the curse. In these studies analyzed in Wellman et al.’s (2001) meta-analysis, knowing that the candy has been eaten or baked may still be enough of a curse to contaminate children’s false belief reasoning.

In light of research demonstrating that the curse of knowledge is stronger in children aged three than in older children (Birch & Bloom, 2003), younger children’s performance on false belief tasks would be more compromised than older children. Thus, traditional false belief tasks are an unfair assessment of any developmental differences in understanding the mind because they will either mask young children’s true competencies or exacerbate their difficulties.

This research supports the claim that there is a fundamental continuity between children’s and adults’ ability to reason about the mind. For children, the heightened potency of the curse leads to more blatant errors in mental state attribution. But even for adults, their own knowledge can substantially contaminate their ability to predict what someone else will believe, know, and do.

References