Believe it, or Not? 
Explaining why children fail the standard false belief task

It is not widely discussed, especially among philosophers of cognitive science, that children before the age of four can pass simpler versions of False Belief Tasks. There has been little discussion (and no consensus) about how to characterize the understanding these younger children manifest in these tasks. Success on these tasks, on the face of it, need not trouble the orthodox interpretation of the Standard False Belief Task (SFBT); these children simply understand the representational nature of belief, and hence, master the full-blown concept of belief, at an earlier age than commonly thought. Recent results have shown, however, that children as young as 13 months of age can pass such simplified tasks, and I think there are good reasons not to attribute mastery of the full-blown concept of belief to children at this age. Based on this evidence, I will argue that the abilities of these young children provide a serious challenge to the orthodox interpretation of the SFBT, and that we need a different analysis of the mindreading abilities of children at all of these ages. The major change that allows the child to pass the SFBT is not, I will claim, understanding the representational nature of belief. I propose an alternative analysis of the developmental data, stressing that understanding beliefs should be distinguished from mastering the full-blown concept of belief, and that the latter may involve capacities that go well beyond what has been described traditionally as aspects of ToM.
Believe it, or Not?
Explaining why children fail the standard false belief task

§1 Introduction

Before a child turns four, he or she seems to have a particularly glaring deficit in understanding certain aspects of belief. Despite his or her relative ease in handling social situations, the child seems to be unaware that people can have false beliefs. Most cognitive scientists working on Theory of Mind (ToM) diagnose this deficit, as measured by an inability to pass the Standard False Belief Task (SFBT), as a failure to understand the representational nature of beliefs. It is not until the child understands the representational nature of belief that he or she comes to master the full-blown concept of belief.

It is not widely discussed, especially among philosophers of cognitive science, that children before the age of four can pass simpler versions of False Belief Tasks. There has been little discussion about how to characterize the understanding these younger children manifest in these tasks. Success on these tasks need not trouble the orthodox interpretation of the SFBT; these children simply understand the representational nature of belief, and hence, master the full-blown concept of belief, at an earlier age than commonly thought. Recent results have shown, however, that children as young as 13 months of age can pass such simplified tasks, and I think there are good reasons not to attribute mastery of the full-blown concept of belief to children at this age. Based on this evidence, I will argue that the abilities of these young children provide a serious challenge to the orthodox interpretation of the SFBT, and that we need a different analysis of the mindreading abilities of children at all of these ages. The revolution that allows the child to pass the SFBT is not, I will claim, understanding the representational
nature of belief. I propose an alternative analysis of the developmental data, stressing that understanding beliefs should be distinguished from mastering the full-blown concept of belief, and that the latter may involve capacities that go well beyond what has been described traditionally as aspects of ToM.

§ 2 A Developing Understanding of Belief

The alternative to the orthodox interpretation of the SFBT is to find blame elsewhere in the mind. Many cognitive scientists have noted that the SFBT places an extremely high cognitive load on children, thus making it a poor candidate for testing ToM in young children (e.g. Bloom and German 2000). If the task demands are lessened (e.g. making the questions simpler, giving memory aids, etc.), many children can pass a version of the task at around the age of three. If these results are legitimate, then the failure of three year olds to pass the SFBT is not adequate evidence that they fail to understand false belief, since the failure was not due to a deficit in ToM. I want to suggest that there is sufficient evidence to show that we need to accept some version of this alternative interpretation.

Following Amanda Woodward’s (1998) work on goal-understanding in infants, Kristine Onishi and Renee Baillargeon (2005) devised a nonverbal version of the false belief task that 15 month-olds can pass. Utilizing the preferential looking paradigm for testing the expectations of preverbal infants, the researchers hoped to show that the children would spend more time looking at a scenario in which their expectations about where the adult would look for an object were violated. If the children understood that seeing leads to knowing, and that the events that they witness will not be available for
someone who does not witness them, and hence, can have no impact on the behavior of the adult, they will find it surprising if an adult puts an object into one container, and is not privy to the fact that the object has been moved into another container, but nevertheless reaches for the object in the new container. On the other hand, if the child lacks an understanding of false belief, as the data for the SFBT suggests, the child should not be surprised when an adult who did not see the switch reaches to the object’s current location. If the child does not realize that her beliefs can be different from another’s beliefs, she should assume that since she saw the object being moved, the adult will know its current location as well.

When the 15 month-olds witnessed a switch in the object’s location, to which the adult did not have access, the children looked substantially longer when the adult reached for the object’s current location than they did when the adult reached for the original location.iii Onishi and Baillargeon take this to suggest that the children expected the adult to reach for the container in which the adult placed the object, and that those expectations were violated, and hence, were of more interest and drew more attention from these children. They argue that this shows that the children understand false belief, even at this extremely young age.iv

More recently, Luca Surian, Stefania Caldi, and Dan Sperber (ms) have gathered evidence that 13 month-olds can also manifest this sort of understanding. In their experiments, children are shown animated video sequences in which a creature searches for food in a room containing different sorts of screens (occluders). During the familiarization stages, the creature watches as a hand places an apple behind one screen and some cheese behind the other. The creature then consistently moves behind the same
screen to obtain the same sort of food (this continues in every episode). In the test trials, the food locations are switched before the creature enters the scene. Two conditions are measured—in one, the screens are low enough for the creature to see the food behind them (the seeing condition); in the other, the screens occlude the food (the not-seeing condition). Surian et alia wanted to measure the children’s expectations about when the creature would be successful in obtaining its favorite food in the novel location. If the children expected the creature to need to see the new location of the food in order to obtain it, they should be surprised (and hence look longer) when the creature goes directly to the new location of the food in the not-seeing condition, but not in the seeing condition. In fact, the data suggest that this is exactly what happened. In the seeing condition, children were not surprised when the creature succeeded in going to its favorite food in its new location, but were surprised when it went to the old location (at which the other food was visible). In the not-seeing condition, the children failed to show any statistically significant expectation about where the creature would go (seemingly attributing ignorance to the creature).

In order to further examine the extent to which children of this age were capable of taking an agent’s perspective into account, Surian et alia changed the experiment slightly. After the same familiarization trials, two new conditions were introduced. In both conditions, the screens fully occluded the food behind them, but in one condition, the creature was present when the hand placed the food in the new locations (the knowing condition), whereas in the other, the creature was not present when the food was placed in the new location (the not-knowing condition). The experiment was designed to determine how the children’s expectations about the two groups might differ despite the
fact that neither group had current perceptual access to what was behind the screens. It could be that the children would have no expectations and would treat the groups similarly; or the children could use past information about where the favorite food had been so many times before; or they could expect the creatures who saw the food placed in the novel location to get the favorite food and show surprise if the creature went to where the food usually was located. If the children are taking perspective into account, Surian et alia expected the children would not show any surprise for the not-knowing condition (since they are as ignorant as in the not-seeing condition) but should be surprised if the creature fails to use the new information about the location of its favorite food. In fact, this is exactly what happened. There was no statistically significant difference in looking times in the not-knowing condition, but the children looked significantly longer when the creature went to the old location in the knowing condition.

Given the intense focus that has been placed on the inability of children to pass false belief tasks before the age of four, these results are remarkable. Children seem to be able to distinguish the agent’s perspective from their own. When an agent has not seen a change to a scene, the children do not expect the agent to act as though it has seen the change. When an agent has seen a change to a scene, children are surprised when the agent fails to take that information into account. When an agent has no information about a scene, children fail to be surprised no matter which choice the agent makes.

§3 Unpacking ‘Representational’

As aforementioned, the most common interpretation of the skills needed to pass SFBT is a representational ToM, one that views beliefs as representations of reality that need not
accurately represent reality and will lead to certain behaviors that would not occur if other representational states were present. This interest in the representational nature of belief stems most clearly from the Theory Theory (TT) account of ToM advanced by Josef Perner, Alison Gopnik, Henry Wellman and others. According to the TT, the child first develops a non-representational conception of both belief and desire. This theory changes into one with a representational conception of desire, but still maintains a non-representational conception of belief, what Perner calls a prelief. Finally, as they mature in their third year, they gain a theory with a representational conception of belief and are able to pass the SFBT.

Shaun Nichols and Stephen Stich (2003) criticize the TT account of the development of the representational conception of desire by arguing that the evidence that TT cites as evidence for the early presence of an understanding of desires as non-representational actually suggests that the desires are understood representationally. I will make a parallel point about belief. The evidence that TT and most theorists cite for the claim that a non-representational conception of belief is used by children until around the time they can pass the SFBT is flawed. In what follows, I will catalog the most common depictions offered by the Theory Theorists for understanding a representational mental state. I will then explain why the tasks summarized above seem to meet these depictions.

Nichols and Stich note that the notion of representation has a murky history, that there is little consensus about how this is to be understood, and describe TT’s attempts to define the notion as “dark pronouncements” (2003, 113). Nevertheless, I think some consensus does emerge. A minimal account would say that a psychological state $\psi$ is
understood representationally when an agent understands that a state of affairs, $S$, can be
\[ \psi \ \text{by} \ A \ \text{even if it is not} \ \psi \ \text{by} \ B. \]
Perner has offered similar accounts. He writes, “a false belief is characterized not only by its being false but also by the holder of the belief deeming it true” (1988, 150). A more telling quotation is as follows:

In contrast, 2-year-olds cannot understand that something that is false can be mistaken as realis, which is exactly what is required for understanding false belief in distinction to pretense, or a sincere but mistaken statement based on a false belief as opposed to a pretend statement. Only around 4 years of age, when they understand perspectival relativity (as a core feature of representations), can children understand that within a different perspective, a proposition that is evidently false from the child’s perspective (i.e. not an appropriate description of reality) can be realis. (Perner et alia 2005, 222)

These depictions all seem to involve Perspectival Relativity—the idea that agents can have different takes, perspectives, or stances towards the world; that these perspectives can be wrong; and that a mistaken perspective can manifest itself in behavior. I take this Perspectival Relativity criterion to be the key element for understanding mental states representationally.

If this Perspectival Relativity is all that is needed to make the conceptual revolution from understanding preliefs to understanding (representational) beliefs, then I believe the experiments described in §2 suggest that children have a representational understanding of belief at as early as 13 months of age, and perhaps even earlier. In these experiments, the children appear to be able to incorporate the beliefs of the agent into expectations about what that agent will do. When the agent’s beliefs should not reflect reality, the child is surprised that the behavior reflects reality. When the agent’s beliefs should reflect reality, the child is surprised when the behavior fails to reflect reality. When the agent should lack the crucial beliefs about reality, the child is not surprised by either outcome. If the children were incapable of understanding perspectives as relative
to agents, the children should have either expected the agent to look where the object was most frequently, or looked to where the object actually was from the child’s perspective (as evidence from the SFBT suggests). However, the children did neither of these things. The children are not attributing preliefs; they are attributing beliefs.

We need a different perspective regarding the role of belief in early ToM. We need to explain what changes in children between 13 months and four years of age, when it comes to belief. There are at least three basic sorts of explanations: (a) ones that describe changes in ToM, (b) ones that describe changes in other cognitive systems, and (c) ones that describe changes in both. Regarding the first sort, TT will have to explain some development in the theory about how minds work, Simulationists will have to explain the development of novel simulating capacities, and hybrid theorists will have to explain the development of either theory, simulating capacities, or both.

I do not see how such moves by TT or Simulationists would help here. Each alternative requires that what the child understood about the mind was somehow defective, that he was not really attributing beliefs because he did not realize that beliefs must have $X$, or that the sort of simulation he did was substantially different than the simulation at age four, because the child’s simulation lacked some feature $Y$. On the face of it, however, there appears to be nothing wrong with the predictions made by the child, based on the understanding of beliefs as representational. Either the child simulated that the agent would look in one place, but the agent did not, causing surprise; or the child expected the agent’s representation to lead to its typical behavioral manifestation, but the actual behavior was unexpected. Without an independently motivated story about what’s wrong with the child’s understanding of how false belief leads to behavior, without
plausible candidates for $X$ and $Y$, I don’t see any reason to think that this sort of understanding of beliefs, in and of itself, changes at around four years of age.

The hybrid theorist has the best *prima facie* explanation available for the ToM-only proposal. If both sorts of capacities are available, then perhaps the child used the wrong one in the early task, and realizes by age four that he should have used the other. Or, if only one sort of ToM understanding was available, it is the availability of the new sort that allows success in the SFBT. The problem is that the hybrid theorist owes us some explanation for why either theory or simulation can be used to pass the nonverbal task, but not the SFBT. Since it seems that adults can use either method in reasoning about false beliefs, either our phenomenology in such tasks is radically misleading, or the failure of the child’s theory/simulating ability remains to be explained. Hence, in explaining what sorts of changes occur to the theory or to simulation to get the child to succeed in SFBT, the hybrid theorist takes on the problems of the monolithic versions of the theory described above$^{xi}$

**§4 Listening to ToM**

I do not have much hope that a ToM-only solution will be able to explain the development adequately. I think the problem is not that ToM, and specifically, the understanding of belief, is deficient; I think the problem is that the rest of the mind is not properly connected to ToM. Hence, I think we need to look outside of ToM to figure out what sort of development is required.

I offered alternatives (b) and (c) as two ways of describing the developmental trajectory of the children, but I think that such a contrast between ToM and the rest of the
mind begs too many questions about modularity. Hence, I will assume that modules, if they exist, are substantially leaky, and reformulate the alternatives as positions along a continuum with very little ToM development (and relevant cognitive development) on one end and very substantial ToM development (and irrelevant cognitive development) on the other.\textsuperscript{xii} I expect an explanation of ToM in children that minimizes ToM development (in its understanding of belief, more precisely) and stresses cognitive development. Specifically, I suggest that it is the improvement of Executive Functioning (EF), Working Memory, and Linguistic Ability,\textsuperscript{xiii} along with a relatively stable understanding of belief, which finally allows children to pass SFBT. I hope that no cognitive scientist would disagree that these cognitive factors are essential to passing SFBT; my aim in this paper is simply to deny what has been assumed by many cognitive scientists—that a misunderstanding of the representational nature of belief is the culprit in failing the task. When we say that at approximately four years of age, children finally master the full blown concept of belief, we are not saying anything specifically new about ToM, but rather something about the interactions between ToM and other cognitive systems and the new abilities the child now has at her disposal in virtue of these interactions.

Defending this proposal in detail is beyond the scope of this paper; the identification of where on the continuum the proper account of development is to be found is an empirical issue. Luckily, several cognitive scientists have begun to explore these issues, and I will give a brief example of the sort of interaction between ToM and other cognitive abilities that exemplifies this explanation of development.\textsuperscript{xiv}
One aspect of EF that has been linked to ToM is inhibitory control (since one must inhibit one’s own beliefs in order to predict the ways contrary beliefs will affect someone else’s behavior). Longitudinal studies have demonstrated that early success in inhibitory control is an excellent indicator of future success in passing ToM tasks." Two general accounts about this relationship have been generated, one that seeks to explain how inhibitory control and other EF factors affect the ToM’s ability to enter into task completion, and another which sees EF as a precursor to ToM development. The problem with these studies is that they equate ToM ability with the ability to pass the SFBT, so there are many questions as to the exact role for ToM here. It could be that EF is an important element in passing SFBT or it could be that EF itself is too meager at the right ages to serve the role of precursor to ToM, especially those early-emerging aspects of ToM.

The existence of the early understanding of belief, then, offers many new paths to explore. For instance, EF could play a specific role within ToM, say in the generation of discrepant beliefs (in Nichols and Stich’s terminology, what gets put into the Possible Worlds Box), where the child can generate some discrepant beliefs at an early age, but due to poor inhibitory control, non-discrepant beliefs will often dominate. Alternatively, poor inhibitory control could play a role in how the information coming from ToM is processed. It could be that ToM is functioning properly, but that beliefs about the current environment are not being as inhibited as the information from ToM (explaining the apparent greater influence of a current state of affairs over counterfactual states of affairs).
In conclusion, the cognitive development of children is a complicated affair. I have offered evidence that despite common opinion, children master the representational nature of belief long before they can pass SFBT. These children understand belief, despite their inability to show it in many contexts. With this fact in mind, it is the difficult job of cognitive science to explain why the children fail to make this understanding of belief evident—why they understand beliefs, yet fail to master them. I think that answering these questions will help us understand what ToM amounts to, and doing so will force us to look at the interrelations among many cognitive systems.

Works Cited


Dennett, D.C. 1978: Beliefs about beliefs. *Behavioral and Brain Sciences*, 1, 568-70.


I am assuming familiarity with this task and the basic issues surrounding it. Current interest in studying ToM capacities can be traced to a paper by David Premack and Guy Woodruff (1978) on the social understanding of chimpanzees. Premack and Woodruff coined the capacity to impute mental states to the self and other agents as ‘Theory of Mind’ and argued that the chimpanzee behavior could only be explained if the chimpanzees utilized a ToM. Daniel Dennett (1978) was dubious about the idea that these chimpanzees really understood mental states like beliefs, and reasoned that it should be possible to detect whether a creature had beliefs about the beliefs of others as opposed to some more limited ability. Heinz Wimmer and Josef Perner (1983) devised a task which they believed would test this capacity of having beliefs about beliefs. This task required understanding that an agent can have a different belief about a situation than the attributor has. This procedure and its variants have come to be known as the false belief task, and it has widely been accepted as the litmus test for ToM. A popular version of the task goes as follows:

A child is shown a series of events involving puppets and some sort of salient object, in this case, a roll of Smarties (the British version of M&Ms). After becoming familiar with the puppets and objects, the test begins with one of the puppets, Sally, coming into the scene, placing the candy into a covered container, and then leaving the scene. In Sally’s absence, another puppet enters the scene, removes the candy from where Sally placed it and places it in a second covered container. The child is asked to report where Sally will look for the candy.

The reasoning behind the experiment is that someone who fully understands the concept of belief must grasp the dependence of belief on observation. Someone who understands the dependence of belief on observation will realize that since Sally did not see the second puppet change the location of the candy, Sally will think that the candy is in the location where she placed it. Such a person will realize that Sally’s belief is different from her own belief because of Sally’s absence. Someone who lacks this understanding will not realize that Sally will not know that the location has changed and will report that Sally will look at the current location of the candy. This ability to predict that Sally will look where she believes the candy to be, even though the child knows that the location of the candy is different, is said to be critical for a full-blown ToM. As aforementioned, this standard version of the task is reliably passed by normal children around the age of four.

Debate about ToM often revolves around whether these abilities are the result of using a tacit theory (the Theory Theory, see Gopnik and Wellman citations) or the result of simulation accomplished by running one’s own practical reasoning system offline (the Simulation theory, see Goldman and Gordon citations). Newer debates stress the need for both sorts of strategies, and are hybrid views. I think the best version of a hybrid view is Nichols and Stich (2003), and this is the basic framework I adopt, making various adjustments based on what I say below. When it matters to stress the differences among these theories, I will make note. Otherwise, I believe that The orthodox view of SFBT failure is shared by members in each of the three camps, whether it concerns belief as something like a scientific concept or simulating with false beliefs, or a combination of both.

In this task, children become familiar with the activity of an adult who spends her time placing objects in covered boxes and retrieving those objects at a later time. In between the placing of the object and the removal of the object, in some cases, a sliding occluding device blocks the adult from seeing the boxes. In some of these occluded cases, the object is moved from one covered box to another covered box. Hence, the children are familiarized with all of the events that will be involved in a false belief experiment.
The time was longer than they ordinarily look when the adult succeeds at getting the object—they had become bored with this scenario.

Josef Perner and Ted Ruffman (2005a, b) have responded to these experiments by offering possible low-level strategies or mechanisms that explain the children’s expectation without appealing to false beliefs. Alan Leslie (2005), Cshiba and Southgate (2006), and Surian et alia (ms) offer replies to these attempts.

See Wellman (1990) and Gopnik and Wellman (1992, 1994).

See Eric Schwitzgebel (1999) for more analysis of the TT account of desire.

As I note below, this isn’t just a problem for TT, but for anyone who explains the failure of SFBT by the failure to understand the representational nature of belief in theorizing or simulating.

This is similar to one of the formulations that Nichols and Stich pull out of the TT discussion (2003, 112). They attribute it to Gopnik and Wellman.

It is true that Perner lists this as a core feature of representations, so maybe this condition is not singularly sufficient. But, TT consistently treats this as the crucial feature when discussing the representational nature of mental states, so even if my claim is too bold, TT still needs to explain why the crucial feature of representations is understood so long before the age of 4.

Surely ToM develops over time, and the child becomes more adept at prediction, but the issue is what this development has to do with understanding beliefs, and whether some different understanding of beliefs, plus a bunch more ToM ability, or merely more or less the same understanding of beliefs, plus a bunch more ToM ability is what explains success in passing SFBT. I think the latter claim is on the right track, but with the addition of other cognitive factors (see below).

I am not denying the obvious fact that there is enormous cognitive development going on between 15 months and 4 years of age; I am only allowing for the possibility that this enormous cognitive development, along with the ToM relatively unchanged from 15 months, is not sufficient for passing SFBT—that a substantially new understanding of belief is required.

Inter alia; this list is not intended to be exclusive.

I do not take my claim to be the trivial assertion that children would be unable to pass SFBT unless they had intact executive functioning, sufficient verbal abilities, and sufficient working memory, since the task is verbal, requires that the child remembers the scenario, and requires that the child pay attention to the scenario in question (and not lose focus). All that is true, and the SFBT tests for these abilities in pretrial examinations. I want to argue that there are much more complicated and interesting relationships between ToM and these cognitive abilities that, when explored, can hopefully explain the success and failure of children in the tasks under discussion.

See Moses et alia (2005) for details.