What does the so-called False Belief Task actually check?

Hanoch Ben-Yami,[1] Maya Ben-Yami[2] and Yotham Ben-Yami[3]

ABSTRACT. There is currently a theoretical tension between young children's failure in False Belief Tasks (FBTs) and their success in a variety of other tasks that also seem to require the ability to ascribe false beliefs to agents. We try to explain this tension by the hypothesis that in the FBT, children think they are asked what the agent *should* do in the circumstances and not what the agent *will* do. We explain why this hypothesis is plausible. We examined the hypothesis in two experiments, each involving a new task. In the first task, the hypothesised misunderstanding of the question leads to failure *without* the need to ascribe a false belief, and we show that failure in this new task is correlated with failure in the FBT. In the second task, passing which requires ascribing a false belief to an agent, and for which we have partial yet encouraging results, the children are asked a question which is unlikely to be misunderstood. Children pass this task much more often than they do a standard FBT. The mentioned tension is thus resolved. We conclude that the so-called False Belief Task probably does not check the ability to ascribe false beliefs but rather linguistic development.

KEYWORDS. False Belief Task; Developmental Paradox; Language Acquisition.

Foreword

In the early 2010s, we developed the hypothesis that the children who fail the False Belief Task (FBT) fail because they misunderstand the question, not because they do not understand that others might be wrong. We started running experiments to confirm it in April 2013, and presented our work-in-progress at a talk at the ESPP 2014 conference (September 16–19, Noto, Sicily), twice as posters, and once as a TED Talk.[4] By then, we'd successfully concluded one of our two experiments and were working on the second. However, for several reasons this work was terminated before its conclusion. Administrative work, studies and other commitments slowed us down considerably; we had difficulties finding Hungarian kindergartens willing to participate; and MBY moved from Budapest to Cambridge, and as young Hungarian children are used to female nurses, work with them became more difficult, many did not cooperate, and the results became less trustworthy. We were also displeased with some features of the second experiment's video, but we could not reshoot it: a better video might yield better results.

However, people familiar with our work have been interested in a more detailed report, some referred to our research in their publications, and the second experiment has been used as a paradigm by some researchers. We therefore thought it is worthwhile to make our partial results available. As we haven't been following the literature closely for the last couple of years, this might make our paper lacking in this respect; still, we simply don't have the time to catch up with recent developments. Because of all this, we decided to make the paper available without trying to publish it in a journal.

---

[1] Corresponding author. Department of Philosophy, Central European University, benyamih@ceu.edu
[2] Institute of Astronomy, University of Cambridge
[3] Department of Music, King's College London
[4] Posters: Budapest CEU Conference on Cognitive Development 2015 (BCCCD 2015), January 8–10; 'The Nature and Origins of Human Cognition', The Berlin School of Mind and Brain, Humboldt University, June 18–19, 2015. TED talk, in Hungarian: https://www.youtube.com/watch?v=gRNZTPZnGlc.

We link below to the videos of our two experiments. It is possible that the videos will need to be moved at a later stage to a different platform. Links to the videos will then be available on HBY's homepage.

1 Introduction

It has been firmly established that most children up to the age of four fail in what is known as the False Belief Task (FBT) (Wimmer & Perner, 1983; Baron-Cohen *et al.*, 1985; Wellman *et al.*, 2001). Namely, most children up to that age, when asked where the agent in the FBT will look, instead of saying that he will look where they should know he *wrongly believes* a desired object is, say he will look where that object *in fact* is. This result has been taken to show that the children who fail the task lack adequate understanding of the concept of belief, and presumably of other related psychological concepts as well. In particular, at least in circumstances similar to those of the FBT, they cannot understand that others might entertain a wrong belief, namely be wrong. Let us call this way of explaining the failure in the FBT, *the cognitive hypothesis*.

The cognitive hypothesis has been challenged by a series of experiments over the last fifteen years, which have shown by means of looking patterns and other nonverbal responses that children who fail to *reply* correctly in the FBT still have the correct *expectations* (Onishi & Baillargeon, 2005; Surian *et al.*, 2007; Southgate *et al.*, 2007; Baillargeon *et al.*, 2010; Kovács *et al.*, 2010; the earliest experiment that might have documented this is Clements & Perner, 1994). In these experiments, an agent has a false belief concerning the location of a desired object, a false belief induced in a similar way to the one induced in typical FBTs. The children then exhibit indications of surprise when the agent goes to where, unbeknownst to him, the desired object in fact is, but not when he goes to where he wrongly thinks the object is; or, they anticipatorily look at the place to which the agent will go if he holds a false belief, and not at the place where the desired object in fact is; and so on (most of these results are reviewed in Baillargeon *et al.*, 2010). These indications of correct expectations have been shown to exist already before the age of one year (Kovács *et al.*, 2010).

Moreover, in an experimental setup in which an agent requested help by referring with either a common noun or pronoun to where he wrongly believed an object is, the children, seventeen month old, took the agent to mean not what was, unbeknownst to the agent, in the place he was referring to, but what the agent wrongly believed to be there, and responded appropriately (Southgate *et al.*, 2010). This experiment has thus shown that the use of complex linguistic behaviour and the need to respond to it do not, as such, block the child's ability to ascribe a false belief and to respond accordingly. Helping behaviour that indicates the attribution to an agent of a false belief was also shown with 18 month old and possibly 16 month old children, this time without complex language use by the agent (Buttelmann *et al.*, 2009; Buttelmann *et al.*, 2014). Even if the looking-pattern and violation-of-expectations experiments of Baillargeon, Kovács and others that were mentioned in the previous paragraph might be explainable by means of an associative response and not as indicating the child's considered view, it is hard to see how these helping-paradigm experiments that involve elicited response can be explained without attributing to the child the ability to ascribe a false belief to others.

The tension between the results is obvious: how come children have the right expectations and understanding, as displayed by their non-linguistic behaviour, yet give the wrong answer? This phenomenon has therefore come to be known as *the Developmental Paradox*. Some attempts to reconcile these results have been made, but we don't find them satisfactory.

One idea, which makes use of an approach that had been suggested before the more recent experiments were made, is that children do have an automatic, innate ability

to ascribe false beliefs, but that they fail the FBT because they lack efficient inhibitory abilities, and in the FBT they cannot inhibit themselves from ascribing to the agent the true belief which they themselves hold (Leslie *et al.,* 2004; Baillargeon *et al*., 2010; He *et al.,* 2012). – But then it is unclear in what sense children *do* have this ability to ascribe false beliefs, if they cannot exercise it; or, if they do exercise it but only as long as they are not *asked* a question that requires inhibitions, as is shown by their having correct expectations, it is unclear why a question should make these inhibitions fail. Moreover, the claim that the ascription to an agent of a belief different from yours requires any inhibitions at all is itself problematic, as it is unclear whether one is in fact inclined to ascribe one's own beliefs to others. So this approach seems unpromising as a framework within which to explain the results of the conflicting experiments.

Another explanation that has been suggested (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; see also De Bruin & Newen, 2012) is that humans have two systems that track beliefs. One is cognitively efficient but limited and inflexible, and is available already to infants; in virtue of this system infants and toddlers have the correct expectations described above, which are conditional on the ascription of false beliefs to agents. The other system, which is highly flexible but cognitively inefficient, develops only later and gradually, and is related to the development of language; its early unavailability is responsible to infants' and toddlers' failures in the FBT. – However, the fact that children who fail the FBT do pass complex tasks that involve nontrivial use of language and the ascription of a false belief (Southgate *et al.,* 2010) make this explanation implausible. Children understood that although an agent, while introducing a name, gestured to a place where one object was, the agent meant to name another object, which unbeknownst to him was removed from that place. This demonstrates both flexibility in reasoning and developed linguistic skills, in contrast to the hypothesis made in (Apperly & Butterfill, 2009) and (Butterfill & Apperly, 2013). We also find the evidence adduced for the existence of such two systems inconclusive.

Both suggestions deserve a more detailed discussion (which is partly found in (Jacob, 2013); see also Heyes's article (2014) and Scott & Baillargeon's response (2014)). Here we develop a different explanation, which we have also tested experimentally.

2 The Misunderstanding Hypothesis

The conflicting results suggested to us that the children who fail the FBT might *understand the situation* but *misunderstand what they are asked*. In the FBT, the children are asked a question in the future tense. Wimmer and Perner asked, 'Where *will* Maxi look for the chocolate?' (1983: 108); Baron-Cohen and his co-researchers asked, 'Where *will* Sally look for her marble?' (1985: 41); and so on. Our hypothesis was that the children think that they are asked by the experimenter not what the agent *will* do, but what the agent *should* do. Namely, that the children take themselves to have been asked, what the best or right action for the agent is.

We find this hypothesis prima facie plausible. First, the circumstances in which an adult asks a young child something to which the adult obviously knows the right answer are commonly educational, the adult trying to teach the child what is the best or right thing to do in the situation and expecting the child to reply accordingly. The *pragmatics* of the FBT situation could therefore incline the child towards understanding the question normatively.

Secondly, when we provide a child with several optional courses of action and want to check whether he knows what he *should* do in the circumstances, we often ask him, what *will* he do? For instance, 'In which box *will* you put the ball', 'Which toy *are you going* to choose', and so on; that is, when asking about the best or correct course of action for the child to take, we use the future tense and not a verb form with the normative

'should' or some equivalent special grammatical construction. So the young child is often asked questions regarding a future action in which the future tense has a normative meaning and not a descriptive one. In addition, children typically master the use of pronouns only in the course of their third year of life. Until then adults often address them by their name, not as 'you' but as 'Jack' and 'Jill'. When asked a normative question, for instance where *should* they put something, they are therefore often asked, 'Where will Jill put the ball?' etc. In this way they are often exposed to the third-person future tense questions in a normative sense. By contrast, it seems that when we ask a young child what another agent will do, we rarely if ever do that when the agent has less information than the child about the situation. Accordingly, our interrogative use of the future tense as applied to a third person in interactions with a young child probably does not allow the child to distinguish the descriptive from the normative, while when we ask the child what *he* will do, the meaning is normative. If the young child generalises from the uses to which he has been exposed, he will reasonably take the question, 'What will *x* do?', in a normative sense. So there are also good *semantic* reasons for the child to misunderstand the FBT question.

Accordingly, it would not be surprising if a young child understands the FBT question, 'Where *will* the agent go?' as asking where the agent *should* go. The child has good reasons for understanding the question in this way. Given this misunderstanding, the child answers the question correctly: since Sally, say, is interested in her marble, she *should* go to where the marble *is*, and not to where she thinks it is. On this interpretation of the results, the children's failure in the FBT does not result from an inability to understand that others might be wrong, but from misunderstanding what they are asked. Let's call this explanation, *the misunderstanding hypothesis*. This hypothesis also explains why the children who fail the FBT pass the tasks which do not involve asking them a question, as in these tasks no linguistic misunderstanding arises. The Developmental Paradox is thus resolved by the misunderstanding hypothesis.

A different hypothesis, which we consider at this place to contrast it with ours, has been suggested by Pierre Jacob (2013) (see also Helming *et al.,* 2014). Jacob suggested that children have a cooperative bias, and that 'being asked a question by the experimenter is likely to trigger [this] bias, whereby young children inappropriately extend a second-person cooperative perspective, requested by the experimenter's communicative action, onto the mistaken agent's instrumental action'. Consequently, they inappropriately turn the agent's motivation into a shared goal, and the distinction between the prediction question meant by the experimenter and the normative question is blurred. The children's 'spontaneous ability to track the mistaken agent's epistemic perspective on the object's location' is in this way disrupted.

Unlike Jacob, we do not assume any cooperative bias. In addition and relatedly, we do not assume any failure of processing by the children who fail the FBT, while Jacob thinks that they fail 'because they are not yet able to maintain a third-person perspective onto an instrumental action while taking a second-person perspective onto the experimenter's communicative action.' Jacob thinks that the activation of a communicative bias motivates the children 'to advise the mistaken agent', and that the children's 'propensity to take this second-person perspective inappropriately spills over onto the mistaken agent's instrumental action, which impairs their performance.' Yet even granting such a cooperative bias, it is hard to see why a question should confuse the children and make them attempt to help not only the experimenter but a third agent as well. By contrast, we think that the children only misunderstand what they are asked, namely, that they think they are asked not what the agent *will* do but what it *should* do.

As can be seen, our hypothesis is more parsimonious than Jacob's. Both hypotheses, however, explain the results of the experiments described above. As will be

seen, both also explain the results of the new experiments we conducted. Additional work, for instance on cooperative biases and their misapplications or on language understanding and acquisition, will be necessary to decide between these hypotheses.

An important contrast between our attempt to explain the Developmental Paradox and other existing attempts is as follows. All other approaches explain the paradox by claiming that the children who fail the FBT are cognitively underdeveloped in some way. We explain it not by hypothesising any such cognitive limitation but by claiming that their exposure to language *justly* makes them misinterpret the question.

We proceed to describe the two experiments we conducted to test our hypothesis. The first should show that the children who fail the FBT fail also a task in which they are asked a similar question but passing which involves no ascription of a mistake to any agent. This would show that our misunderstanding hypothesis explains not only the failure in the FBT, but other, related failures as well, which the cognitive hypothesis cannot explain. The second experiment should show that children pass an FBT when the question is simple to understand, and in particular does not require understanding of the use of the future tense. This would show that cognitive hypothesis cannot explain why children fail the standard FBT.

3 First Experiment: The Bad Bear Task

We tried to devise a task that will establish the misunderstanding of the question by the children. To this end, we tried to devise a task in which:

1. passing the task would not require ascribing a false belief to any agent
2. knowing what the agent *will* do is very simple even to the child
3. yet if the child would respond by saying what is *best* for the agent to do, he would fail the task

Condition (1) ensures that if children fail the task, this is not due to an inability to ascribe a false belief. The other conditions would support the hypothesis that if the children fail, they fail because of the suggested misunderstanding of what they are asked. If failure is further correlated with failure in the FBT, this would support the hypothesis that both failures are due to the same misunderstanding by the children of what they are asked.

The experiment consisted of three different tasks, and we tried to put to each child all three tasks.

In the first, 'Squirrel Task', which constitutes our baseline condition and was intended to rule out any floor effect, the children watch a video in which the opening frame shows two screens, a red one to the left and a green one to the right, with a gap between them. A squirrel puppet then comes out at the back between the two screens and says, in Hungarian, 'Hello! I'm Hanna, I'm a squirrel. I'm going to go out now and stand in front of the red screen.' The video is then frozen, and the child is asked where will Hanna the squirrel go, to the red screen or to the green one? We made sure before playing the video, when the children saw only the opening frame, that they knew which screen is red and which green. We considered passing the Squirrel Task a necessary condition for minimal understanding of the language and the situation, and only children that answered it correctly were taken into account when considering the results of the next two tasks.

The second, 'Bad Bear Task' is the crucial one. It opens with the same two-screen frame. A bear comes out from behind between the screens, says that he is Gergelymel the bad bear, and that he beats cute little puppets. He next says that he is going to hide in front of the red screen, as he then does, and the child sees him standing in front of the red screen. Next, a much smaller puppet comes out from behind between

the screens, from where she cannot see the hiding bear, and says, using almost the same words that the squirrel used, 'Hello! I'm Kovági, a cute little puppet. I'll go out now and stand in front of the red screen'. The picture is then frozen and the child is asked, again in the same way as in the Squirrel Task, where will Kovági the little puppet go, to the red screen or to the green one?

We also wanted to check whether there is any correlation between the results of the Bad Bear Task (BBT) and those of the FBT. We therefore filmed the following version of the FBT, the 'Princess Task': a 'princess' puppet comes out behind the same two screens holding a chocolate and says, 'Mmm... My chocolate is delicious! I'll put the chocolate now in front of the red screen, leave, and I'll be back to eat it.' After putting the chocolate in front of the screen and leaving, a mouse comes in, moves the chocolate in front of the green screen, and leaves. The princess then returns, standing in the gap behind the screens so that she cannot see where the chocolate is, and says, 'I'm back for the chocolate'. The picture is again frozen and the child is asked in the same way, where will the princess go. Notice that the child *wasn't* asked, where the princess will *look for the chocolate*, nor was the chocolate mentioned at all by the experimenter. This latter question might be misinterpreted as a question about where the princess will *look and see*, or where she will *find* the chocolate; it might thus make it more difficult to pass the task. We instead asked a more neutral question, one which is also identical to the one asked in the two other cases, namely, where will the princess go.

In order to pass the BBT the child need not ascribe any false belief to the puppet Kovági. Kovági says where she will go, and if the child understands her simple sentence, an understanding supported by passing the Squirrel Task, then he should simply report what the puppet said. However, if the child thinks that he is asked what Kovági *should* do (i.e. what is best for her to do) then he will say that she will go to the green screen, despite her explicit assertion that she will go to the red one. Accordingly, passing the BBT does not involve the ability to ascribe a false belief, while failing it supports the hypothesis that the child thinks he is asked what the agent should do.

The question whether failures in the BBT and FBT are correlated is significant for the following reason. If they are, then the factors responsible for failure in the BBT are active in the FBT as well. These factors, as explained above, do not involve an inability to ascribe a false belief. If they can also bring about failure in the FBT, they will consequently make redundant the hypothesis that the failure in the FBT is due to an inability to ascribe a false belief. We may then still maintain that the child *can* ascribe a false belief, and that his failure in the FBT is due to these other factors. The child's ability to ascribe a false belief will then explain why he does have the correct expectations and understanding, as demonstrated in the publications mentioned above. The tension between the failure in the FBT and having the correct expectations and understanding will then be resolved.

Since we did get good correlations between failure in the BBT and failure in the FBT, we shall discuss here a question that might arise concerning our considerations above. In the video, Kovági is *ignorant* of the presence of the bad bear in front of the red screen. Now, passing the BBT does not require that the child ascribe a false belief to the agent. Yet, although ignorance is not a false belief, if a child cannot ascribe a false belief to an agent because he ascribes only true beliefs (correlating thought and reality), one might claim that the child would equally be incapable of ascribing ignorance (which breaks this correlation). So if passing the BBT requires the ascription of ignorance to the agent, one might suspect that the BBT is not *that* different from the FBT and that failure in both is due to the same inability.

There is good empirical evidence that already 12 month old children can ascribe ignorance to agents (Knudsen & Liszkowski, 2013). But in any case, the ascription of

ignorance in the BBT plays a different role from the ascription of a false belief in the FBT. Inability to ascribe ignorance in the BBT would help *passing* the task, while an inability to ascribe false belief in the FBT would help *failing* in the task. To see this, suppose that the only misleading factor in the BBT is the assumed inability to ascribe ignorance. So in the BBT the child thinks that Kovági *knows* the bad bear is in front of the red screen and that she said she will go there while knowing that. If so, the child has no reason for thinking Kovági will go to the green screen. Accordingly, if the alleged inability to ascribe ignorance and false belief is the only misleading factor, children who fail the FBT will *pass* the BBT and the failures will not be correlated.

On the other hand, assume the child thinks he should say what is best for the agent, and that he may or may not have the mentioned inability. Now, whether or not Kovági knows the bad bear is in front of the red screen, she better go to the green screen. Accordingly, on this assumption, whether or not the child ascribes to Kovági ignorance, he will fail the BBT. For the same reason he will fail the FBT whether or not he has the mentioned inability. So if the child understands the question as if he is asked what is best for the agent to do, the ascription of the alleged inability to him is redundant.

3.1 Results

The study took place in a quiet room in kindergartens. Usually one of the local teachers was also present and helped with the children. We asked altogether 58 children, ranging in age from just over two years old to just over six years (30 girls, 26 boys, in 2 cases the gender wasn't registered). Of these, a few refused to cooperate in some or all tasks, and in a couple of cases we had technical problems in the execution of the experiment, so we ended with 50 valid results.

With the older children, we very quickly saw that the tasks are very simple for them and we were afraid that they might lose their patience and not watch all the three videos. So while we usually showed the squirrel video first, we showed it to some of them last, with the result that five children didn't stay to watch it. Since the purpose of the squirrel video was to make sure that the children have minimal understanding of the language and the situation, it was practically redundant with the older children. Indeed, all these five children passed both the BBT and the FBT.

The results we got are as follows, showing in the table only the results of those children who did not fail the squirrel task (i.e., the necessary condition for minimal understanding):

All Results
(Failed Squirrel: 7/50)

	Failed Bear Task	Passed Bear Task
Failed Princess Task	14	6
Passed Princess Task	5	18

The results clearly show dependence between the failures in the tasks ($p=0.00148$). Of those who failed the FBT (20 children), 70% also failed the BBT, while of those who passed the FBT (23), only 22% failed the BBT. If we take into account the likely fact that a few children just guessed, lost attention and replied at random, or other 'noise' effects, and that probably these children are divided roughly equally between the four possibilities, we get that the correlation between the failures in case the children reply on the basis of their understanding of the scenarios and of the questions must be even higher.

If we break the results according to age groups, we see the expected progress from failing both tasks to passing both:

Age: 2.0–3.5
(Failed Squirrel Task: 4/21)

	Failed Bear Task	Passed Bear Task
Failed Princess Task	10	2
Passed Princess Task	3	2

Age: 3.5–5.0
(Failed Squirrel Task: 3/21; not asked on Squirrel: 2, both passed both other tasks and are included in the table below)

	Failed Bear Task	Passed Bear Task
Failed Princess Task	4	3
Passed Princess Task	2	9

Age: 5.0–6.5
(Failed Squirrel Task: 0/8; not asked on Squirrel: 3, all passed both other tasks and are included in the table below)

	Failed Bear Task	Passed Bear Task
Failed Princess Task	0	1
Passed Princess Task	0	7

4 Second Experiment: The Falling Screen Task

The second experiment was intended to involve a task the passing of which is conditioned on ascribing a false belief to an agent, but where the question put to the child does not use the future tense or any other phrase which the child might understand as normative and not descriptive. Indeed, the question should be simple, and while relevantly resembling that of standard FBTs, unlikely to be misunderstood by the child.

Something which children are asked from a very early age is, 'Where is *x*': 'Where's mummy?', 'Where's daddy?', 'Where's teddy?', 'Where's light?', and so on. Answering it can be done by mere pointing. Using this question in an FBT would require that, when the child is asked the question, the agent is already at one of the two places, but that the child should have to work out which place it is, and that working it out correctly should involve ascribing a false belief to the agent. We thus shot the following video, the 'Falling Screen Task'.

The video starts with a stage covered by a screen. A puppet enters in front of the screen and, by means of a rope, pulls up the screen with much effort. The puppet twice fails and the screens falls back despite its effort, and only in the third time it succeeds and the screen remains lifted. The puppet then leaves the scene.

On the stage behind the screen we see two boxes, a white one to the left and a black one to the right, otherwise identical. Behind the boxes there's a screen, with an opening in its middle. The 'princess' puppet then enters, carrying a jar of colourful candies. She states that her candies are wonderful, that she'll put them in a box and later come back to collect them. She walks to the left, white box, and one can hear her walking. She lifts its lid, which squeaks when lifted, puts the jar inside the box, closes the lid with a thump, and leaves. One can again hear her walking out.

Once she has left, a 'king' puppet enters from the same place, laughs maliciously, goes to the white box, opens it, takes out the jar and closes the box. The same squeak and thump are heard. He then carries the jar to the other, black box to the right, opens the box – an identical squeak is heard – puts the jar inside, and closes the lid (the same

thump is heard). He then laughs maliciously again, and leaves. – We later came to suspect that the king puppet may have been too frightening and distracting for some children. This may have interfered with the performance of some of the children, who were quite young.

After the king has left, the princess returns, stands at the middle of the stage, and declares that she came back for her candies. At that moment, the curtain drops. We then hear the princess walking behind the screen, and we hear the squeak of a box being opened. At that stage, the video is frozen and the child is asked, *Where is the princess?*

4.1 Results

The study was performed in conditions similar to those of the first study. However, because of much editing, the quality of the video ended up lower than that of the videos of the first, Bad Bear experiment, and this might have made it more difficult for the children to follow. Also, not all rooms were equally quiet, a thing which might have distracted some children and affected the results.

We present results obtained with children aged between two years to three years and six months. We had 32 results which could be considered valid.

Age Range	No. of Subjects	Pass No. (%)	Fail No. (%)
2.01–2.06	5	4 (80%)	1 (20%)
2.07–3.00	9	5 (56%)	4 (44%)
3.01–3.06	18	7 (44%)	11 (56%)
2.01–3.06	**32**	**16 (50%)**	**16 (50%)**

Eight children either did not answer or their response wasn't clear enough to be included (e.g., the direction in which they pointed wasn't clear).

From the third time we ran the experiment on, we first asked the children, 'Where is the princess?', and then, after they answered, 'Where are the candies?'. This is an improvement, as a correct answer to 'Where are the candies?' provides a baseline condition of understanding. This, however, reduces the number of valid results from 32 to 14, too few for statistical significance. Still, the results were,

Age Range	No. of Subjects	Pass No. (%)	Fail No. (%)
2.05–3.00	5	4 (80%)	1 (20%)
3.01–3.06	9	4 (44%)	5 (56%)
2.01–3.06	**14**	**8 (57%)**	**6 (43%)**

Hopefully, improving the video in the ways indicated would provide more reliable results. We still thought that the tendency the results we obtained show makes it worthwhile to include them in the paper.

5 Discussion and Conclusions

Failure in the False Belief Task is correlated with failure in the Bad Bear Task. Accordingly, the factors responsible for failure at the BBT are also available for explaining the failure at the FBT. If our hypothesis is right and children fail the BBT because they think they are asked what is best for the agent to do, then this explains their failure in the FBT as well: what is best for the FBT agent – the princess in our version – is not to go to where she wrongly believes the chocolate is but to go to where the chocolate in fact is. Failure in the FBT is therefore not due to the child's alleged inability

to ascribe a false belief. In fact, the experiments from the recent decade showing that children do have the correct expectations and understanding, as measured by looking time and other responses, support the claim that children from a very early age do ascribe false beliefs to agents and expect the agents to act accordingly. There is no tension between the failure in the FBT and having the correct expectations or other responses. In fact, 'False Belief Task' is a misnomer: it does not check whether the child can ascribe a false belief, since children fail it for other reasons (misunderstanding what a question means), despite their ability to ascribe false beliefs.

The hypothesis of this paper, that children think they are asked what is best for the agent, explains the failure in the Bad Bear Task, the failure in the so-called False Belief Task, the correlation between the two failures, and the fact that the children who fail the FBT still have the correct expectations concerning what the agent will do and even show on other nonverbal tasks understanding of the speaker's intentions despite the speaker's wrong belief (Southgate *et al.,* 2010). It also has, as we noted earlier, *prima facie* plausibility, given (a) the common lack of linguistic distinction between descriptive and normative choice questions when the child himself is asked; and (b) given the fact that when adults ask a young child a question to which they obviously know the answer, they often do it in a pedagogical context, checking whether the child knows what is the *right* thing to do. By contrast, the hypothesis that children fail the FBT because they cannot ascribe a false belief doesn't explain the failure in the BBT or the correlation between the failures, and it is in obvious tension with the children's having the correct expectations etc. For all these reasons, the 'misunderstanding the question' hypothesis is preferable over the 'failure to ascribe a false belief' one for explaining the results of the FBT.

The misunderstanding hypothesis can also explain why deaf children of hearing parents, who typically demonstrate language delays, are also significantly delayed in passing the FBT, while deaf children from deaf families, who typically are not linguistically delayed, perform identically to same-aged hearing children (Schick *et al.*, 2007). The former children obviously understand later than children without language delay, be they deaf or hearing, the question asked in the FBT.

Although the results of the Falling Screen Task (FST) cannot be taken as conclusive, they do support the misunderstanding hypothesis. If children of the age 2.01 to 3.06 couldn't indeed understand that agents might be wrong, as the cognitive hypothesis has it, then they should have failed the FST as often as they fail the FBT. This was clearly not the case, as about half the children passed the FST. Moreover, that about half the children answered correctly shows that a simple FBT question, unlikely to be misunderstood, improves performance. It thus supports the claim that in standard FBTs the children fail to understand the question, not the situation. The still low percentage of success, slightly above 50%, might be a result of the relative complexity of the task and the issues with the video and execution of the experiments mentioned above.

According to the present hypothesis, the thing that normally changes around the age of four is not that the child acquires the ability to ascribe a false belief. What the child acquires around this age is a more nuanced understanding of questions. He learns to distinguish between *will* and *should* as applied to other agents, between cases in which he is expected to say simply what is going to happen and cases in which he is expected to show that he knows what would be the best thing to do. – It is left for future research to examine whether this hypothesis about the acquisition of a more nuanced linguistic understanding can be supported by data from additional kinds of linguistic behaviour.[5]

[5] An alternative explanation of the experiments mentioned in the Introduction was suggested by Rubio-Fernández and Geurts, an explanation supported by their own results (Rubio-Fernández & Geurts, 2013). Rubio-Fernández and Geurts claimed that nonverbal false belief tasks impose weaker linguistic demands (if

If our considerations in this paper are correct, then at present there are no good reasons for claiming that children who fail the so-called False Belief Task cannot ascribe false beliefs to agents, namely, cannot understand that others might be wrong.

Acknowledgments

We thank Judit Lonyai-Sagiv and Zita Zsemberovszky for their help with the children in three kindergartens; Hagit Glickman for her help with the statistics; Szilvia Takacs for translating the consent forms into Hungarian; Bence Tóth for his help with editing the videos; and the members and visitors of the CEU Department of Cognitive Science for many valuable discussions.

any) compared to verbal ones, and unlike the latter they are normally designed to minimally interfere with children's natural ability to track events from the agent's point of view. For this reason, they claimed, children who fail the verbal tasks, being unable to follow the agent's point of view, often do pass the simpler nonverbal tasks. We think, however, that this cannot explain why children fail the BBT, in which they simply need to accept the report of Kovági on where she will go, as they did in the case of the squirrel. Moreover, children who failed the standard FBT still passed the task of Southgate *et al.* (2010), which involved complex interaction between intended reference, actual reference and false belief. Lastly, to the extent that the results of the FST can be trusted, children can pass a modified FBT which involves linguistic demands. So Rubio-Fernández and Geurts's explanation seems problematic.

Rubio-Fernández and Geurts also conducted an alternative FBT experiment, in which the children were drawn by the experimenter into her attempt to deceive a Duplo girl doll, by moving an object the doll was interested in to a new location while the doll could not see what is going on. The children were then asked what will happen next or where will the doll go, and they were given the opportunity to continue the play with the doll themselves. When the element of deception was present, a much higher percentage of children passed the task (RF&G's Experiment 1, by contrast to their Experiment 2a). Rubio-Fernández and Geurts claimed that this was so because it was easier for the children to track the doll's perspective during Experiment 1. However, as children exhibit the correct anticipatory behaviour in these and more complex scenarios, as described above, this explanation has to make the problematic assumption that once asked even a simple question the children get confused and lose track of what the agent knows. In contrast, we explain the difference in the results of the experiments as follows. In Experiment 1, having been drawn into the attempt to deceive the doll, the children do not take the opportunity given to them to continue playing with the doll as a request to show what is best for the doll; after all, they are part of the deceptive enterprise. When they are not part of a deception attempt (Experiment 2) they do tend to think they are expected to show what is best for the agent, and they therefore fail the task. Accordingly, it seems that our misunderstanding hypothesis can explain the results obtained by Rubio-Fernández and Geurts, as well as additional results that pose a problem to their explanation.

References

Apperly, I. A. & Butterfill, S. A. 2009. Do humans have two systems to track beliefs and belief-like states? *Psychological Review* 116(4): 953–970. doi: 10.1037/a0016923.

Baillargeon, R., Scott, R. M., & He, Z. 2010. False-belief understanding in infants. *Trends in Cognitive Sciences* 14: 110–118. doi: 10.1016/j.tics.2009.12.006.

Baron-Cohen, S., Leslie, A. M. & Frith U. 1985. Does the autistic child have a 'theory of mind'? *Cognition* 21(1): 37–46. doi: 10.1016/0010-0277(85)90022-8.

Buttelmann, D., Carpenter, M. & Tomasello, M. 2009. Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* 112: 337–342. doi: 10.1016/j.cognition.2009.05.006.

Buttelmann, D. Over, H., Carpenter, M., & Tomasello, M. 2014. Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of Experimental Child Psychology* 119: 120–126. doi: 10.1016/j.jecp.2013.10.002

Butterfill, S. A. & Apperly, I. A. 2013. How to Construct a Minimal Theory of Mind. *Mind & Language* 28(5): 606–637. doi: 10.1111/mila.12036

Clements, W. A. & Perner, J. 1994. Implicit understanding of belief. *Cognitive Development* 9: 377–395. doi: 10.1016/0885-2014(94)90012-4

De Bruin, L. C. & Newen, A. 2012. An association account of false belief understanding. *Cognition* 123: 240–259. doi: 10.1016/j.cognition.2011.12.016.

He, Z., Bolz, M. & Baillargeon, R. 2012. 2.5-year-olds succeed at a verbal anticipatory-looking false-belief task. *British Journal of Developmental Psychology* 30, 14–29. doi: 10.1111/j.2044-835X.2011.02070.x.

Helming K. A., Strickland B. & Jacob P. 2014. Making sense of early false-belief understanding. *Trends in Cognitive Sciences* 18(4): 167–70. doi: 10.1016/j.tics.2014.01.005.

Heyes, C. M. 2014. False belief in infancy: A fresh look. *Developmental Science* 17(5): 647–659.

Jacob P. 2013. Solving the puzzle about early belief-ascription. Unpublished manuscript.

Knudsen, B. & Liszkowski, U. 2013. One-Year-Olds Warn Others About Negative Action Outcomes. *Journal of Cognition and Development* 14(3): 424-436. doi: 10.1080/15248372.2012.689387.

Kovács, Á. M., Teglas, E. & Endress, A. D. 2010. The social sense: Susceptibility to others' beliefs in human infants and adults. *Science* 330: 1830–1834. doi: 10.1126/science.1190792.

Leslie, A. M., Friedman, O. & German, T. P. 2004. Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences* 8: 528–533. doi: 10.1016/j.tics.2004.10.001.

Onishi, K. H. & Baillargeon, R. 2005. Do 15-month-old infants understand false beliefs? *Science* 308: 255–258. doi: 10.1126/science.1107621.

Rubio-Fernández, P. & Geurts, B. 2013. How to pass the false-belief task before your fourth birthday. *Psychological Science* 24: 27–33. doi: 10.1177/0956797612447819.

Schick, B., de Villiers, P., de Villiers, J. & Hoffmeister, R. 2007. Language and Theory of Mind: A Study of Deaf Children. *Child Development* 78(2): 376–396. doi: 10.1111/j.1467-8624.2007.01004.x

Scott, R. M. & Baillargeon, R. 2014. How Fresh a Look? A reply to Heyes. *Developmental Science* 17(5): 660–664.

Southgate V, Chevallier C & Csibra G. 2010. Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science* 13(6): 907–912. doi: 10.1111/j.1467-7687.2009.00946.x.

Southgate, V., Senju, A. & G. Csibra 2007. Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science* 18: 587–592. doi: 10.1111/j.1467-9280.2007.01944.x

Surian, L., Caldi, S. & Sperber, D. 2007. Attribution of beliefs by 13-month-old infants. *Psychological Science* 18: 580–586. doi: 10.1111/j.1467-9280.2007.01943.x

Wellman, H. M., Cross, D., & Watson, J. 2001. Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development* 72: 655–684. doi: 10.1111/1467-8624.00304

Wimmer, H. & Perner, J. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13(1): 103–128. doi: 10.1016/0010-0277(83)90004-5.