COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 41 (2017) 218–241

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12331

Can We Forget What We Know in a False-Belief Task? An Investigation of the True-Belief Default

Paula Rubio-Fernández

Centre for the Study of Mind in Nature, University of Oslo

Received 14 July 2014; received in revised form 1 September 2015; accepted 10 September 2015

Abstract

It has been generally assumed in the Theory of Mind literature of the past 30 years that young children fail standard false-belief tasks because they attribute their own knowledge to the protagonist (what Leslie and colleagues called a "true-belief default"). Contrary to the traditional view, we have recently proposed that the children's bias is task induced. This alternative view was supported by studies showing that 3 year olds are able to pass a false-belief task that allows them to focus on the protagonist, without drawing their attention to the target object in the test phase. For a more accurate comparison of these two accounts, the present study tested the true-belief default with adults. Four experiments measuring eye movements and response inhibition revealed that (a) adults do not have an automatic tendency to respond to the false-belief question according to their own knowledge and (b) the true-belief response need not be inhibited in order to correctly predict the protagonist's actions. The positive results observed in the control conditions confirm the accuracy of the various measures used. I conclude that the results of this study undermine the true-belief default view and those models that posit mechanisms of response inhibition in false-belief reasoning. Alternatively, the present study with adults and recent studies with children suggest that participants' focus of attention in false-belief tasks may be key to their performance.

Keywords: Theory of Mind; Sally–Anne task; True-belief bias; Competition for attention; Eye tracking

1. Introduction

Thirty years of Theory of Mind research have repeatedly shown that children under 4 years of age fail standard false-belief tasks (Wellman, Cross, & Watson, 2001). In the Sally–Anne task, for example, Sally puts a marble in a box before going out to play. During her absence, Anne moves the marble to a basket, setting the scene for the false-belief

Correspondence should be sent to Paula Rubio-Fernández, Centre for the Study of Mind in Nature, Blindernveien 31, Georg Morgenstiernes Hus, 0313 Oslo, Norway. E-mail: paula.rubio-fernandez@ifikk.uio.no

(FB) question: "When Sally comes back, where will she look for her marble?" (Baron-Cohen, Leslie, & Frith, 1985). Young children err by predicting that Sally will look for her marble in the basket, rather than in the box where she left it.

Young children's performance in FB tasks has been interpreted as evidence of a true-belief (TB) bias in early social reasoning, according to which children are predisposed to attribute their own beliefs to others and predict their behavior on the basis of their own knowledge (see, e.g., Baillargeon, Scott, & He, 2010; Carlson & Moses, 2001; Fodor, 1992; Goldman, 2006; Leslie, Friedman, & German, 2004; Leslie & Polizzi, 1998; Nichols & Stich, 2003; Perner, Lang, & Kloo, 2002). We have recently proposed an alternative to the received view: Young children fail standard FB tasks because they lose track of the protagonist's perspective, not because they attribute their own knowledge to the protagonist by default. In our view, various task manipulations may disrupt the process of perspective tracking in young children (see Rubio-Fernández & Geurts, 2013, for discussion). The form of the FB question is critical in this respect, as it draws children's attention to the target object, thus increasing the salience of the wrong response (e.g., "Where does Sally think the marble is?").

We tested this hypothesis in two developmental studies, which confirmed our predictions. When 3 year olds performed a standard FB task designed to allow young children to focus on the protagonist throughout the narrative (a task we named the "Duplo task"), they made the correct prediction 80% of the time. However, this performance was dependent on the test question not mentioning the object (e.g., "What happens next? What is the girl going to do now?" used in Rubio-Fernández and Geurts [2013]; also "Where will Lola go now?" as Lola is standing between the two containers, used in Rubio-Fernández and Geurts [2015]). By contrast, children performed below chance level when they were asked a standard FB question ("Where will the girl look for her bananas?"; Rubio-Fernández & Geurts, 2013). Likewise, 3 year olds failed the Duplo task when the target object was mentioned in the test phase (Rubio-Fernández & Geurts, 2015), either in a control question ("Where are the bananas now? [...] What happens next? What is Lola going to do now?") or to make the protagonist's goal explicit ("Now Lola is very hungry and wants a banana. What happens next? What is Lola going to do now?").

In parallel with our developmental work, I have recently tested the role of participants' attention focus in an eye-tracking study with adults. In two change-of-location FB tasks (with and without a test question) adults were able to correctly anticipate the protagonist's actions without suffering interference from their own knowledge. However, adults showed an early preference for the actual location of the object when the containers momentarily disappeared from the scene during the test phase. These results were interpreted as supporting the hypothesis that perspective tracking is a continuous process that is dependent on attentional resources and can therefore be disrupted by task manipulations, even in adults (Rubio-Fernández, 2013). More specifically, the results of this experiment are in line with the view that the two containers in a change-of-location FB task represent two different perspectives on the location of the object: the outdated (corresponding with the protagonist's) and the updated (corresponding with the participant's). Since both representations of the object compete for attention during processing (Altmann

& Kamide, 2009), adults momentarily fell back on their own perspective when their focus on the protagonist's was disrupted by the sudden disappearance of the containers.

2. Performance factors in false-belief tasks

Wellman et al.'s (2001) meta-analysis of FB tasks revealed that the relative salience of the target object in the test phase was a reliable performance factor in studies with preschoolers. This finding hinged on the results of early studies showing that children under 4 years of age perform at chance (rather than below chance) in unknown-location tasks in which they know that the target object has been moved, but they ignore to where exactly (Bartsch, 1996; Wimmer & Perner, 1983). The results of Rubio-Fernández and Geurts (2013, 2015) further suggest that children's focus of attention in the test phase (whether it is on the object or on the protagonist) is an important predictor of 3 year olds' performance in standard FB tasks. Children's focus of attention at the time of answering an FB question may also shed light on young children's poor performance in unexpected contents FB tasks.

For example, in the Smarties task (Perner, Frith, Leslie, & Leekam, 1989), children are shown a tube of Smarties. Children guess that it contains chocolates, but then the Experimenter shows them that the tube, in fact, contains pencils. Children are then asked a control question, "What's in here?" and then the test question "When the next kid comes in and we ask him what's in here, what is he going to say?" As in the control question condition by Rubio-Fernández and Geurts (2015), young children tend to give the same response to both questions, namely the actual contents of the tube. This is also the case when the test question concerns their previous belief about the contents of the box (Gopnik & Astington, 1988). This response pattern could be related to the type of perseverance errors that have been found with the Appearance/Reality task, in which 3 and 4 year olds tend to give the same answer to two different questions (Deák & Enright, 2006).

In those unexpected contents tasks that do not include a control question prior to the test question (e.g., Carlson & Moses, 2001), the wrong response may nonetheless be relatively salient since the Experimenter always corrects the child's false belief about the contents of the tube before asking him the test question. Supporting this hypothesis, early studies by Lacohée and colleagues suggest that reducing the salience of the wrong response in unexpected contents tasks by physically representing both responses in the experimental setting improves performance in younger groups (Freeman & Lacohée, 1995; Mitchell & Lacohée, 1991).

While children's focus of attention in FB tasks seems to be an important factor in 3 year olds' performance, this factor alone does not explain all aspects of young children's performance in Theory of Mind tasks. Wellman and colleagues, for example, have developed a scale of Theory of Mind tasks that reveals a consistent developmental progression in the preschool years. In this scale, children who are able to pass a standard FB task are normally able to pass a "diverse-belief task" (in which they need to understand that two people can have different beliefs about the same object) but not necessarily vice versa (see Peterson, Wellman, & Liu, 2005; Peterson, Wellman, & Slaughter, 2012; Wellman & Liu, 2004). This progression, however, cannot be explained by children's focus of attention in the test phase as similar standard questions were used in both tasks.

Likewise, research on the relationship between language development and Theory of Mind have revealed a correlation between language ability and FB understanding, which is not explained by children's focus of attention in FB tasks (see Low, 2010; Milligan, Astington, & Dack, 2007). It is therefore important to note that the results of Rubio-Fernández and Geurts (2013, 2015) do not invalidate the FB task as a reliable Theory of Mind test, although they suggest that better task designs that control for the relative salience of the wrong response may reveal more insightful results with 3 year olds than previous studies have provided.

3. Differences between the two accounts of early failure in FB tasks

According to the TB-default model put forward by Leslie et al. (see, e.g., Friedman & Leslie, 2004a,b, 2005; Leslie, German, & Polizzi, 2005; Leslie & Polizzi, 1998; Leslie et al., 2004), humans possess an innate Theory of Mind Mechanism that automatically attributes other people relevant beliefs and desires. Leslie et al. (2005: 47–48) argue that TB attribution is an ecologically valid default because people's beliefs are usually true. This means that in cases where another person is mistaken (e.g., in an FB task), the TB default needs to be actively inhibited. Contrary to this model, we argue that it is the design of standard FB tasks that may make children lose track of the protagonist's perspective and fall back on their own. This is particularly problematic for children under 4 years of age, who may lack the necessary executive control to revert to the protagonist's perspective when she returns to the scene (after the object has been moved) or once their attention has been drawn to the target object during the test phase.

Given that both the TB default and the attention-focus accounts assume that young children respond to the FB question from their own perspective and require a certain level of executive control to revert to the protagonist's perspective, it may at first seem that these two accounts are not so different. In what follows, I shall give three arguments showing that, despite these commonalities, there are fundamental differences between the two accounts.

First, the TB-default model assumes that when 3 year olds fail standard FB tasks, they are effectively tracking the protagonist's perspective; they may wrongly attribute their own knowledge to the protagonist, but their response is nonetheless informed by their Theory of Mind. In clear contrast with this assumption, we argue that the standard FB question *disrupts* the process of perspective tracking. That is, 3-year-old children are able to track an agent's perspective provided that they are allowed to focus on the protagonist throughout the FB narrative. However, when children's attention is drawn to the target object in the test phase, their focus on the protagonist is disrupted, which may cause them to respond from their own perspective. This erroneous response, however, does not necessarily involve mental state attribution, contrary to what Leslie, Friedman, and colleagues claim.

Another important difference between the TB-default and the attention-focus accounts is that Leslie, Friedman, and colleagues assume that infants, toddlers, and children are able to mentally represent beliefs, and do so in performing FB tasks (see also Baillargeon et al., 2010). In contrast with this approach, our investigation is only concerned with those attentional factors that could potentially affect perspective tracking in standard FB tasks before age 4. Therefore, we do not assume that infants, toddlers, children, and adults necessarily use the same mental processes and representations in passing FB tasks (see Rubio-Fernández, 2013; Rubio-Fernández & Geurts, 2013, for discussion). For our research, we define perspective tracking as the ability to form expectations about other people's actions on the basis of observations of their behavior; at what age this ability involves mentally representing other people's beliefs is beyond the scope of our investigation. Our starting point is therefore radically different from that of Leslie, Baillargeon, and colleagues.

The third critical difference between the TB-default and the attention-focus accounts regards the role of inhibitory processes in passing standard FB tasks. According to Leslie, Friedman, and colleagues (Friedman & Leslie, 2005; Leslie et al., 2004) and also according to Baillargeon, Scott et al. (Baillargeon et al., 2010; Scott & Baillargeon, 2009), correctly responding to a standard FB question requires inhibiting the prepotent TB response. More specifically, Scott, Baillargeon, Song, and Leslie (2010: 392) claim that "inhibition should predict performance in any false-belief task that involves verbal interaction with an experimenter, regardless of the type of response required." As was recently pointed out by Helming, Strickland, and Jacob (2014), these theories cannot account for the results of Rubio-Fernández and Geurts (2013, 2015) since there is no principled reason why inhibiting the TB default and selecting the correct FB response should be challenging for 3 year olds in the Sally–Anne task, but not in the Duplo task, given that children know where the target object is in both FB tasks.

Contrary to the theories put forward by Leslie, Friedman, and colleagues, on the one hand, and by Baillargeon, Scott et al., on the other (see also Carlson & Moses, 2001), I do not simply assume that passing a standard FB task requires response inhibition. Because I do not take the wrong response to be necessarily dominant, I hypothesize instead that passing a standard FB task is dependent on executive control only to the extent that the participant's focus on the protagonist may be disrupted by task manipulations (e.g., when the protagonist leaves the scene, as in Rubio-Fernández and Geurts [2013; Experiment 2a], or when the containers momentarily disappear from the scene, as in Rubio-Fernández [2013; Experiment 2]. A standard manipulation that may require response inhibition (at least in children) is the mention of the target object in the test phase (see Rubio-Fernández & Geurts, 2015). However, as long as children are allowed to focus on the protagonist throughout the FB narrative, passing a standard FB task need not require inhibition of the erroneous TB response.

Despite these clear theoretical differences, empirically testing the two accounts of young children's failure in standard FB tasks may not be possible since both accounts assume that 3-year-old children respond to the FB question from their own perspective (albeit for different reasons) and require a certain level of executive control to revert to the protagonist's perspective. What should be possible, however, is to test the TB-default model with adults, and that was indeed the aim of this study.

4. Testing the TB default with adults

The TB-default model is related to a more general phenomenon known as the "curse of knowledge" (Camerer, Loewenstein, & Weber, 1989): People suffer interference from their own perspective when estimating other people's knowledge, opinions, or impressions. For example, people tend to overestimate the amount of general consensus on opinions they themselves hold, or they have the mistaken impression that something is simple just because they are familiar with it (see Nickerson, 2001, for a review). The curse of knowledge has also been investigated in FB reasoning in adults, with positive results (see Birch & Bloom, 2007; Coburn et al., 2015; cf. Ryskin & Brown-Schmidt, 2014; Sommerville, Bernstein, & Meltzoff, 2013).

All the Theory of Mind studies of the curse of knowledge have used off-line tasks that required participants to make an estimate of a certain outcome. Crucially, these designs did not tap the earliest stages of processing and were therefore not an accurate test of the TB-default model proposed by Leslie and colleagues. In this sense, the results of these studies are compatible with the attention-focus view that our own perspective is in competition with that of another when we are trying to adopt their perspective (Altmann & Kamide, 2009; Rubio-Fernández, 2013), without necessarily supporting the view that we start by automatically attributing our own beliefs to others. An accurate test of these two models therefore requires using on-line tasks that tap the earliest stages of FB reasoning and can reveal any automatic tendency that adults might have to respond to the FB question from their own perspective. The eye-tracking measures used in the present study aimed to provide such a test.

I take the following to be the chief tenets of the TB-default model: (a) people automatically attribute their own beliefs to others, and so in situations where another person is mistaken, (b) the TB default needs to be inhibited in order to accurately predict the mistaken person's actions (Leslie et al., 2004, 2005). Wang and Leslie (unpublished data) specifically argue that "the [TB] bias arises from theory-of-mind competence itself and takes the form of a *rational prior* to attribute one's own belief to others" (first page; emphasis added). Importantly, these authors claim that eye gaze can tap this rational prior from childhood to adulthood (ibid.).

Given the two tenets outlined above, I take FB tasks to be a critical test of the TB-default model since their design may reveal (a') any initial tendency to respond to the test question according to the participant's own knowledge and (b') whether this automatic tendency needs to be inhibited in order to correctly predict the agent's behavior. Since the TB default is argued to be a rational prior in FB reasoning, sufficiently fine-grained measures of processing should reveal evidence of both (a') and (b') if the TB-default model is correct.

Continuous eye-tracking measures, such as those used by Rubio-Fernández (2013), and response times as an on-line measure of inhibition (Neill & Westberry, 1987) offer a promising methodology for accurately testing the specific predictions of this model.

Experiments 1a and 1b aimed to test the first tenet of Leslie's TB-default model and investigated whether people have an automatic tendency to respond to the FB question according to their own knowledge. Experiment 2 and a follow-up experiment aimed to test the second tenet of Leslie, Friedman et al.'s account (see also Carlson & Moses, 2001; Scott & Baillargeon, 2009; Scott et al., 2010) and investigated whether the TB response needs to be inhibited in order to accurately predict the protagonist's actions.

5. Experiment 1a

In a recent eye-tracking study, Rubio-Fernández (2013) presented adult participants with one of two computer versions of the Sally–Anne task: a standard task including an FB question, and a narrative version that did not have a test question at the end of the story. As mentioned in the Introduction, in one of the experiments of the study the two containers momentarily disappeared from the scene in the test phase, which resulted in a disruption of the process of perspective tracking: In both the standard and the narrative versions of the Sally–Anne task, adult participants showed an initial preference for the current location of the object when the containers reappeared in the scene (contrary to what had been observed in the normal, un-interrupted versions of the task). The aim of Experiment 1a was to further investigate the effect of this visual disruption; more specifically, whether the initial preference for the current location of the object was the result of a simple disruption of perspective tracking (as argued by Rubio-Fernández, 2013) or whether it reflected an underlying TB default.

Eye-tracking studies with adults using the blank-screen paradigm have revealed processes of spatial indexing in memory retrieval (see Hoover & Richardson, 2008; Johansson & Johansson, 2013; Richardson & Spivey, 2000). For example, Altmann and Kamide (2009) presented participants with short stories while they looked at a static scene depicting a story event. The scene was then removed and participants' eye movements were monitored while the story continued. Even if participants were just "looking at nothing" (as Ferreira, Apel, and Henderson [2008] refer to this paradigm), eye-tracking measures revealed that participants were fixating on the same regions of the screen that were previously occupied by the objects that were now being mentioned.

If a short blanking of the screen in an FB task reveals a similar process, the resulting disruption of perspective tracking could be related to an underlying TB default in memory retrieval. If that is the case, the early fixations on the location of the object should also be observable in subsequent trials since spatial indexing occurs even after participants habituate to the blank screen. On the other hand, if Rubio-Fernández (2013) correctly interpreted the original results and the sudden disappearance of the containers simply distracts participants momentarily, then the effect should not be observable in a second trial. These alternative hypotheses were tested in the first experiment.

5.1. Method

5.1.1. Participants

Here 48 undergraduate students at Princeton University took part in the experiment. They were all native speakers of English and participated for monetary compensation.

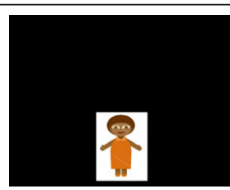
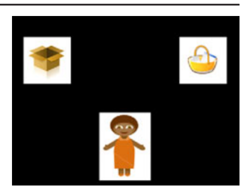
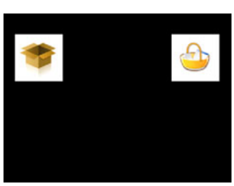
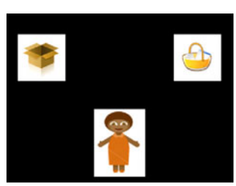
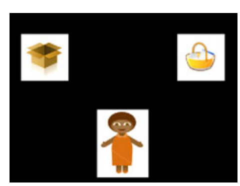
5.1.2. Materials and procedure

The FB task was an extended computer version of the classic Sally–Anne task in which the two containers momentarily disappeared from the scene during the test phase. Since the background of the slides was black, the disappearance of the containers effectively resulted in a blank screen that lasted for 2,000 ms while the narrative continued (see Table 1).

In the cartoon, Sally and Anne were two friends who went to kindergarten together. Each girl had a favorite toy and a container in which she left it when going home at the end of the day. Participants were familiarized with the setting of the story in two warm-up trials (i.e., “Where is Anne’s box? And where is Sally’s basket?”) before they were tested in one of three experimental conditions:

1. *TB Condition/Single Trial:* Sally tries to put her doll in her basket, but since her basket is getting full, she decides to put it into Anne’s box instead.
2. *FB Condition/Single Trial:* Sally puts her doll in her basket and goes home. While she is away, Anne moves Sally’s doll from Sally’s basket to her box.
3. *FB Condition/Second Trial:* The FB condition described above was also presented in a second trial. Other than for the order of presentation, this FB condition was identical to the first one. The first trial in the FB/Second Trial condition was a TB trial featuring Anne and her horse (since Anne’s box is getting full, Anne decides

Table 1
Cartoon slides and text corresponding with the test phase in Experiments 1a and 1b

Experiment 1a			
Experiment 1b			
Question	*When Sally comes back,*	*where will she*	*look for her doll?*

to put her horse into Sally's basket). At the end of the TB trial, participants were asked "Where is Anne's horse now?" The point of this first TB trial was to familiarize participants with the brief disappearance of the containers during the test phase (so that by the second trial they would know that the containers reappeared during the question) while asking participants a different test question (so that they would not simply anticipate the FB question in the second trial).

In all three conditions participants were asked: "When Sally comes back [Sally re-appears], where will she look for her doll?" Participants were evenly and randomly allocated to one of the three experimental conditions. Each participant was therefore tested in one critical trial (TB/Single Trial, FB/Single Trial, or FB/Second Trial) as is normally done in developmental Theory of Mind studies.

The point at which the containers reappeared on the screen was synchronized with the onset of the verb "look for" and was taken as the onset for eye fixations in order to make an accurate measure of the initial stages of perspective taking. A fixation was defined as an eye movement that remained on one of the regions of interest for a minimum of 100 ms.

Eye movements were recorded with an infrared eye-tracking system (504 Pan/Tilt; Applied Science Laboratories Inc., Bedford, MA) that measured eye position at a rate of 60 Hz. The eye-tracking system had a resolution of 0.14 degrees and could detect differences in relative eye position of ~0.25 degrees. Participants were seated in a comfortable chair and their heads were secured in a chin rest for the duration of the experiment (~4 min).

Participants were given standard instructions that described their role as a control group in an experiment aimed at the children. Participants watched the Sally–Anne cartoon on a computer screen and listened to the accompanying story. At various points in the story, they were asked a comprehension question, which they answered using one of two labeled keys on the computer keyboard. The response keys were parallel to the two containers on the screen and participants had to press them with their dominant hand.

5.1.3. Measures and predictions

Four measures of processing were collected: (i) accuracy of first fixation (i.e., whether participants first fixated on the correct or the incorrect target), (ii) latency of first accurate fixation, (iii) proportion of inaccurate fixations, and (iv) response times. Measures (i) and (ii) were established from the onset of the verb "look for" in the question (which is when the containers re-appeared in the scene); measure (iii) covered the segment from the onset of the verb "look for" until the participant responded to the question; and measure (iv) was made from the offset of the question.

I predicted that the results of the FB/Single Trial condition would replicate those observed by Rubio-Fernández (2013) and reveal a disruption of perspective tracking; that is, more inaccurate first-fixations and longer first-accurate-fixations than in the TB/Single Trial condition. The key question was whether similar results would be observed in the FB/Second Trial condition. According to Rubio-Fernández (2013), participants would be momentarily distracted by the sudden disappearance of the containers in the first trial, but they should not be surprised that the containers disappeared again in the second trial, hence performing more accurately. On the other hand, if the blanking of the screen reveals spatial indexing, this effect should be replicable across trials, given that spatial indexing occurs even after participants habituate to the blank screen (see Hoover & Richardson, 2008; Johansson & Johansson, 2013; Richardson & Spivey, 2000). This result would support the view that there is a TB default in memory retrieval.

5.2. Results

The proportions of first accurate fixations are plotted in Fig. 1. A chi-square test with Yates' correction revealed a significant difference between the two FB conditions, with more accurate-first-fixations being observed in the second trial, $\chi^2(1, N = 32) = 6.222$, $p = .013$. Relative to the TB control, participants in the FB/Single Trial condition performed significantly worse, $\chi^2(1, N = 32) = 10.800$, $p = .001$; while the difference with the FB/Second Trial condition was not significant, $\chi^2(1, N = 32) = 0.286$, $p = .593$.

The average latencies of the first accurate fixation in each condition are plotted in Fig. 2. A one-way ANOVA revealed a significant effect of condition, $F(2, 47) = 11.043$, $p < .001$. Independent-samples t-tests revealed a significant difference between the two FB conditions, with the fastest first-accurate-fixations being observed in the second trial, $t(30) = 3.186$, $p = .003$. Relative to the TB control, performance in the FB/Single Trial condition was significantly slower, $t(30) = 3.885$, $p = .001$; while the difference with the FB/Second Trial condition was not significant, $t(30) = 1.340$, $p = .190$.

The average proportions of inaccurate fixations in each condition are plotted in Fig. 3. A one-way ANOVA revealed a marginally significant effect of condition, $F(2, 47) = 2.948$, $p = .063$. Independent-samples t-tests revealed a significant difference between the two FB conditions, with the lesser proportion of inaccurate fixations being observed in the

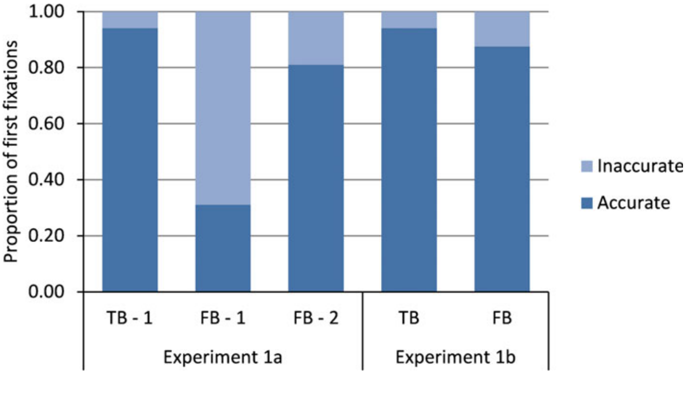


Fig. 1. Proportions of participants who made an accurate or inaccurate first-fixation in Experiment 1a (Conditions 1: unexpected visual disruption; Condition 2: expected visual disruption) and in Experiment 1b (No visual disruption). First fixations were measured from the onset of the verb "look for" in the question.

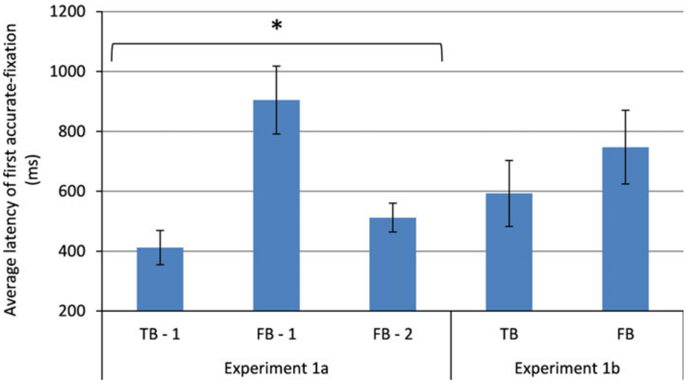


Fig. 2. Average latencies of first fixation on the correct target (in ms) in each condition of Experiments 1a (Visual disruption) and 1b (No visual disruption). First accurate fixations were measured from the onset of the verb "look for" in the question (Standard Error bars; *p < .001).

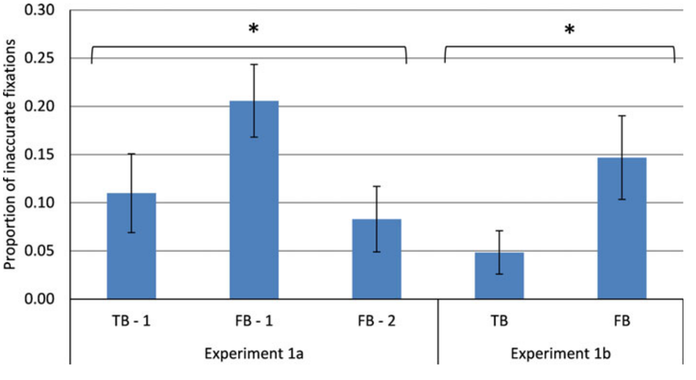


Fig. 3. Average proportions of fixations on the wrong target in each condition of Experiments 1a and 1b. Inaccurate fixations were measured from the onset of the verb "look for" in the question until the participant responded (Standard Error bars; *p < .064).

second trial, $t(30) = 2.415$, $p = .022$. Relative to the TB control, the proportion of inaccurate fixations was greater in the FB/Single Trial condition, albeit the difference only approached significance, $t(30) = 1.737$, $p = .093$; the proportion of inaccurate fixations was smaller in the FB/Second Trial condition, but the difference was not significant, $t(30) = .494$, $p = .625$.

The average response times in each condition are plotted in Fig. 4. A one-way ANOVA revealed a significant effect of condition, $F(2, 47) = 3.501$, $p = .039$. Independent-samples t-tests revealed a significant difference between the two FB conditions, with the fastest response times being observed in the second trial, $t(30) = 2.345$, $p = .026$. Relative to the TB control, performance in the FB/Single Trial condition was significantly slower, $t(30) = 2.126$, $p = .042$; while the difference with the FB/Second Trial condition was not significant, $t(30) = .103$, $p = .919$.

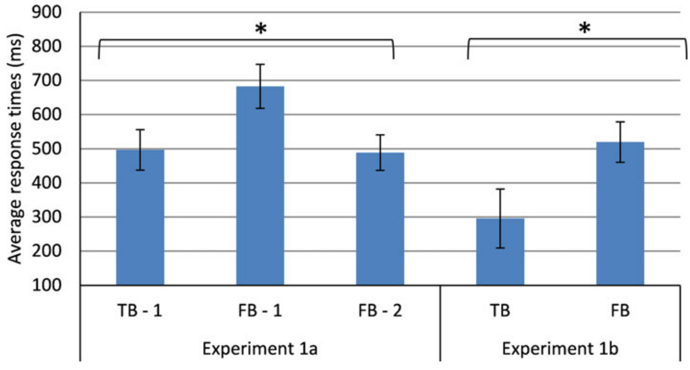


Fig. 4. Average response times (in ms) in each of the conditions of Experiments 1a and 1b. Response times were measured from the offset of the question (Standard Error bars; *$p < .042$).

5.3. Discussion

The results of Experiment 1a confirm that the sudden disappearance of the containers makes participants momentarily lose track of the protagonist's perspective and fall back on their own, as argued by Rubio-Fernández (2013). The fact that the initial preference for the actual location of the object was not observed in a second FB trial (when the disappearance of the containers was not surprising anymore) challenges the alternative interpretation that a brief blanking of the screen reveals a TB default in memory retrieval, since that effect should have been observed across trials.

Another pattern of results that challenges the TB-default model is that the results of the FB/Second Trial condition and the TB control were comparable in all measures. This comparison, however, may not be valid for two reasons. First, since the TB control was administered in a first trial, it is possible that participants' performance was compromised by the sudden disappearance of the containers (i.e., the blanking of the screen might have been disruptive not only in the FB condition, but also in the TB control). Second, performance on the FB/Second Trial condition may have benefited from greater familiarization with the task and practice effects with the questions relative to the single-trial conditions. Given these potential confounds, a more accurate comparison between TB and FB reasoning in adults was made in Experiment 1b. For this purpose, both conditions were administered in a first trial and without a visual disruption of the scene.

6. Experiment 1b

Theory of Mind studies with adults indicate that processing information about another person's false beliefs is more costly than processing information about reality (Apperly, Back, Samson, & France, 2008; Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Back & Apperly, 2010; Birch & Bloom, 2007). However, eye-tracking studies using implicit FB tasks show that adults are able to anticipate a mistaken agent's behavior without suffering an initial interference from their own knowledge (Ferguson, Scheepers, & Sanford, 2010; Rubio-Fernández, 2013; Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Lam, Bayliss, & Dux, 2012; Senju, Southgate, White, & Frith, 2009). None of these eye-tracking studies, however, has compared adults' performance in standard FB and TB tasks, a test that should shed light on the issue of whether people suffer from a TB default in FB reasoning.

According to Leslie and Polizzi, "the calculation of a false belief involves first identifying a true-belief content, followed by the inhibition of that content. Inhibition allows attention to disengage from the true-belief content and move to the alternative, non-factual content of the false belief" (1998: 247; see also Leslie et al., 2005). If this model is correct, eye movements during the processing of an FB question may reveal an initial preference for the actual location of the object, which should then be corrected when the TB default is inhibited. Therefore, replicating the results of the single-trial conditions of Experiment 1a in Experiment 1b (i.e., without a visual disruption of the scene) would support Leslie's TB-default model.

Under the competing assumption that people do not attribute their beliefs to others by default, I do not predict any particular difference between the TB and FB conditions in the early processing measures. As shown by Rubio-Fernández's (2013) eye-tracking study, when adult participants follow an FB narrative, they are tracking the protagonist's perspective and hence anticipate the correct outcome prior to the protagonist's return. Crucially, if adults are allowed to focus their attention on the protagonist uninterruptedly, they are able to anticipate to which container she will return to without suffering interference from their own knowledge of the location of the object.

However, given that in the FB condition, there are two different perspectives on the location of the object (the protagonist's and the participant's) which compete for attention during processing (Altmann & Kamide, 2009), I predict that this competition will result in slower responses in the FB condition relative to the TB control. Unlike the TB-default model, however, my prediction is that the potential tension between the two perspectives should be observable in later processing measures, rather than resulting from an initial tendency to give the wrong response. The results of Rubio-Fernández (2013) offered initial support to the attention-focus account and revealed that the tension between the two perspectives in an FB narrative emerges during the processing of the question (when adult participants revealed a momentary hesitation between the two containers), but no TB bias was observed at the initial stages of processing the question.

Experiment 1b aims to extend the results of Rubio-Fernández (2013) by comparing FB- and TB-reasoning in adults using early and late processing measures. This comparison should be a critical test of the predictions of the TB-default and attention-focus accounts.

6.1. Method

6.1.1. Participants

A different group of 32 undergraduate students at Princeton University took part in the experiment. They were all native speakers of English and participated for monetary compensation.

6.1.2. Materials and procedure

The materials and procedure were the same as in Experiment 1a, with the single exception that the containers remained on the screen throughout the narrative (see Table 1). Participants were familiarized with the setting of the story in two warm-up trials before they were tested in one of two experimental conditions, a TB and an FB Condition (see description of Single Trial conditions in Experiment 1a). Participants were evenly and randomly allocated to one of the two conditions, and hence tested on a single critical trial.

6.1.3. Measures and predictions

The same four measures of processing that were collected in Experiment 1a were collected again in Experiment 1b (i.e., accuracy of first fixation, latency of first accurate fixation, proportion of inaccurate fixations, and response times).

According to the TB-default model, participants in an FB task attribute their true belief to the protagonist by default and need to inhibit it so they can correctly predict the protagonist's behavior. This means that the interference of the TB default should already be observable in the early processing measures (i.e., accuracy of first fixation and latency of first accurate fixation), thus revealing that participants' automatic tendency is to respond according to their own knowledge. By contrast, my prediction is that adult participants should be tracking the protagonist's perspective in an FB narrative (provided that their focus of attention is not disrupted). In this view, the competition for attention between the two perspectives in an FB task should be observable in the later processing measures (i.e., proportion of inaccurate fixations and response times) rather than resulting from an initial preference for the TB response.

6.2. Results

The proportions of first accurate fixations are plotted in Fig. 1. A chi-square test with Yates' correction did not reveal a significant difference between the TB and FB conditions, $\chi^2(1, N = 32) = 0.00$, $p = 1.00$.

The average latencies of the first accurate fixation in each condition are plotted in Fig. 2. An independent-samples t-test did not reveal a significant difference between the TB and FB conditions, $t(30) = 0.935$, $p = .357$.

The average proportions of inaccurate fixations in each condition are plotted in Fig. 3. An independent samples t-test revealed a marginally significant difference between the TB and FB conditions, with the larger proportion of inaccurate fixations being observed in the FB condition, $t(30) = 2.007$, $p = .054$.

The average response times in each condition are plotted in Fig. 4. An independent samples t-tests revealed a significant difference between the TB and the FB conditions, with the fastest response times being observed in the TB control, $t(30) = 2.140$, $p = .041$.

6.3. Discussion

The results of Experiment 1b challenge the TB-default model since adult participants showed a comparable performance on the early processing measures of the TB and FB conditions. That is, participants' eye movements in the FB condition did not show an initial preference for the actual location of the object that needed to be corrected. It is important to note that the expected pattern of results was observed in the FB/Single Trial condition of Experiment 1a (which used an unexpected visual disruption of the scene) and therefore it cannot be argued that the eye-tracking measures used in this study are not fine-grained enough to test the predictions of the TB-default model.

The later processing measures did reveal a difference between TB and FB reasoning in adults. This difference, however, can be explained as a result of the competition for attention between the two representations of the object (the outdated and the updated) which characterizes FB reasoning but does not affect TB reasoning (since the protagonist and the participant have the same perspective). This interpretation of the data is preferable, other things being equal, as it is more parsimonious than assuming that people attribute their beliefs to others by default and need to inhibit the TB default in FB scenarios.

7. Experiment 2

The aim of this experiment was to test the second tenet of the TB-default model: namely, that overcoming the TB default requires an executive process of response inhibition (Leslie et al., 2004, 2005; see also Carlson & Moses, 2001; Scott & Baillargeon, 2009; Scott et al., 2010). In line with Leslie's model, Petrashek and Friedman (2011) showed that in an avoidance-desire task, 3-year-old children inhibited the target that was to be avoided. For example, in one of their stories, a girl wanted to avoid playing under a big red bucket, and children were asked to predict under which bucket she would play (given a choice of three big buckets, including the red one). Next, children were asked to predict under which bucket a boy would play, provided that the boy and the girl did not want to play under the same bucket. The results of this "lingering inhibition task" revealed that 3 year olds continued to avoid the big red bucket when predicting the boy's choice.

Petrashek and Friedman's task did not involve FB reasoning, but Friedman, Leslie, and colleagues have used avoidance-desire FB tasks in earlier investigations of the TB default (see, e.g., Friedman & Leslie, 2004a,b, 2005). Following Friedman and Leslie (2005), Petrashek and Friedman (2011) argued that the effect of avoidance desire on Theory of Mind tasks is not simply an increase in task complexity, and concluded, in view of the lingering inhibition observed in their study, that Theory of Mind processing generally involves inhibition.

Experiment 2 investigated an alternative interpretation of the results of Petrashek and Friedman (2011): The lingering response-inhibition observed by Petrashek and Friedman may not be generally related to Theory of Mind reasoning but to the negation inherent in avoidance (which is a negative desire). It has long been shown that processing negated sentences (e.g., "The door is not open") involves suppressing the positive content of the sentence (Kaup, Lüdtke, & Zwaan, 2006; MacDonald & Just, 1989). Therefore, in order to test whether Theory of Mind reasoning involves inhibition, as these authors claim, I measured lingering inhibition in adults and compared the effect of a negated question ("Where ISN'T Sally's doll?") to the effect of a standard FB question ("Where will Maxi look for his horse?") on the response that is supposed to be suppressed.

7.1. Method

7.1.1. Participants

Here 36 undergraduate students at University College London took part in the study. They were all native speakers of English and participated for monetary compensation.

7.1.2. Materials and procedure

Participants were presented with a cartoon similar to the one used in Experiments 1a and 1b. In this story, Sally and Maxi were two friends in kindergarten, each with a favorite toy and a container to keep it. After participants responded to two warm-up trials (i.e., "Where is Sally's box? And where is Maxi's basket?"), they were presented with the following trials (schematic version):

- Negated-question trial

  Sally and Maxi change places and borrow each other's container to keep their toys.

  Inhibitory question: "Where ISN'T Sally's doll?" (marked stress on the verb)

  Test question: "Where is Sally's doll?" (unmarked stress)

- Control trial 1

  Control question A: "Where does Sally normally keep her doll?"

  Control question B: "Where does Maxi normally keep his horse?"

- FB-question trial

  Maxi puts his horse in his basket and goes home. Before Sally leaves, she takes Maxi's horse out of the basket and puts it in her box.

  Inhibitory question "Where will Maxi look for his horse?"

  Test question: "Where will Sally look for the horse?"

- Control trial 2

  Control question A: "Where does Maxi normally keep his horse?"

  Control question B: "Where does Sally normally keep her doll?"

The order of presentation of the experimental trials (Negated Question/FB Question) and of the test questions relative to the inhibitory questions (before/after) were manipulated, resulting in four lists of materials. Participants were evenly and randomly assigned to one of the four lists.

In both experimental trials, correctly responding to the inhibitory question may require suppressing the incorrect response (e.g., in the Negated-question condition, identifying the container that does not hide Sally's doll may involve suppressing the container that hides it). I was therefore interested in participants' response times to the test questions, as these required accessing the response that would have been suppressed in responding to the inhibitory question. Hence, if responding to the inhibitory questions does involve suppressing the wrong answer, accessing this answer after the inhibitory question should be harder than before the inhibitory question. Participants' response times to the test questions before and after the inhibitory questions were therefore compared in both experimental conditions (manipulated between participants).

As in previous experiments, participants were given standard instructions explaining that they were a control group in a study designed for children. They were going to watch a cartoon on a computer screen and at various points during the narrative they had to respond to a comprehension question using the labeled keys on the computer keyboard. The response keys were parallel to the two containers on the screen and participants had to press them with their dominant hand. It was stressed that participants had to respond as fast and accurately as possible.

7.1.3. Measures and predictions

All four pairs of questions used in this experiment had the same syntactic structure and required alternating answers (i.e., A-B or B-A). This feature of the design should result in facilitation of the second response, other things being equal. I therefore first tried to establish a baseline level of facilitation comparing response times to the control questions in the Control 1 and Control 2 trials. Since the order of the control questions was constant across lists, the baseline comparison was made within participants.

The key test was whether rather than observing facilitation of the second response, the experimental trials revealed inhibition of this response (since it would be the wrong answer to the inhibitory question). If that was the case, responding to the test question should take longer after the inhibitory question than before. Given what has been observed in studies on negation, I predicted that responding to the negated question should involve inhibition of the positive content of the question, hence resulting in longer response times to the test question after the negated question than before. According to Leslie, Friedman, and colleagues, the same pattern of results should be observed in the FB trial, since responding to the FB question involves inhibition of the TB default. Alternatively, if either of the inhibitory questions did not involve suppression of the incorrect response, the results of the test questions should in principle be comparable to those observed in the control condition.

7.2. Results

Two participants had to be replaced because they had responded incorrectly to an inhibitory question (one in each experimental condition). The results of the Control trials provided a baseline level of facilitation, confirming that repeating the syntactic structure of the question and alternating the responses speeds up responding to the second question of a pair (Control 1: Average RT Question A: 1,245 ms [*SD*: 743]; Question B: 487 ms [*SD*: 372]; $t(35) = 7.693$, $p < .001$; Control 2: Average RT Question A: 878 ms [*SD*: 511]; Question B: 460 ms [*SD*: 263]; $t(35) = 6.869$, $p < .001$). The key test for the present investigation was whether the experimental trials would also reveal facilitation (given the repeated syntactic structure of the question and the alternating responses) or suppression of the wrong response to the inhibitory questions.

The average response times to the test questions before and after the inhibitory questions are plotted in Fig. 5. Independent-samples *t*-tests revealed inhibition in the Negated-question condition, with responses to the test question being significantly slower after the inhibitory question than before the inhibitory question, $t(34) = 2.917$, $p = .006$. The opposite pattern of results was observed in the FB condition, where responses to the test question were significantly faster after the critical FB question than before, $t(34) = 3.484$, $p = .001$. No comparison was made between the test question in the Negated-question condition ("Where is Sally's doll?") and the FB condition ("Where will Sally look for the horse?") because the two questions are very different, both in their syntactic form and in the cognitive demands that they pose (the former being factual and the latter requiring a prediction of the protagonist's behavior).

7.3. Discussion

As it was predicted, inhibition of the wrong response was observed in the negated-question condition (corresponding with the positive content of the negated question). By

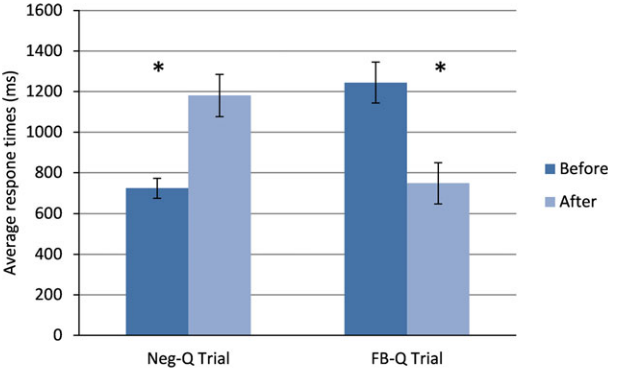


Fig. 5. Average response times (in ms) to the test questions before and after the inhibitory questions (Negated vs. FB). The order of the test questions was manipulated between participants (Standard Error bars; $^*p < .01$).

contrast and contrary to the predictions of Leslie, Friedman, and colleagues, the wrong response to the FB question (corresponding with the participants' true belief) was actually facilitated in the following question (probably because of the repetition of the syntactic structure of the question and the alternating responses, as suggested by the results of the Control condition). The results of Experiment 2 therefore suggest that the lingering inhibition of the avoided target that was observed by Petrashek and Friedman (2011) was due to the negation inherent in avoidance, rather than to inhibition being generally involved in Theory of Mind reasoning, as these authors claimed.

8. Follow-up experiment

To confirm the results of Experiment 2, the experiment was repeated with a new group of 32 undergraduate students from UCL, introducing one variation to the original design: In order to maximize the chances of observing inhibition of the TB response, I used shorter questions in the FB trial ("Where will Maxi look?"/"Where will Sally look?") which were comparable in length to those in the Negated-question condition (<1,000 ms; see Neill & Westberry, 1987).

The results of this follow-up experiment replicated the original findings: Significant facilitation was observed in the Control condition (Control 1: Average RT Question A: 1,135 ms [*SD*: 643]; Question B: 399 ms [*SD*: 202]; $t(31) = 7.229$, $p < .001$; Control 2: Average RT Question A: 873 ms [*SD*: 464]; Question B: 475 ms [*SD*: 208]; $t(31) = 5.578$, $p < .001$), while a significant degree of inhibition was observed in the Negated-question condition (Average RT Before inhibitory question: 703 ms [*SD*: 204]; After: 1,149 ms [*SD*: 429]; $t(30) = 3.807$, $p < .002$). Crucially, the TB response was again facilitated in the FB-question condition (Average RT Before inhibitory question: 1,024 ms [*SD*: 572]; After: 491 ms [*SD*: 192]; $t(30) = 3.529$, $p < .002$). No comparison was made across the Negated- and FB-question conditions as the corresponding test questions were not directly comparable in terms of syntactic structure and cognitive demands.

These results, together with those observed in Experiment 2 pose a serious challenge to the TB-default model, which claims that FB reasoning involves inhibition of the TB default (Leslie et al., 2005; Petrashek & Friedman, 2011; see also Carlson & Moses, 2001; Baillargeon et al., 2010; Scott et al., 2010).

9. General discussion

According to Leslie and colleagues, children and adults attribute their beliefs to others by default. In a standard FB task, it is therefore necessary to inhibit the TB response in order to correctly predict the mistaken character's behavior (Leslie et al., 2005; Scott et al., 2010). The present results undermine this view. First, two eye-tracking experiments using the Sally–Anne task showed that adults were able to anticipate the protagonist's actions without initially suffering interference from their own knowledge. Second, two

FB tasks measuring response inhibition revealed actual facilitation of the TB response. It must be stressed that the measures used in this study were sensitive enough to bear out the predictions of the TB default: Participants' eye movements did reveal an initial preference for the actual location of the object when the visual scene was momentarily disrupted (replicating the results of Rubio-Fernández [2013]). Likewise, in the response-inhibition task, the wrong response to a negative question was indeed inhibited (in line with previous studies on negation; e.g., Kaup et al., 2006).

While the results of the eye-tracking experiments did not reveal an initial TB bias, later measures of processing revealed poorer performance in the FB condition relative to a TB control. However, the hesitation observed in the FB condition is explained by an alternative account, according to which the two containers in a FB task represent the two perspectives on the location of the object: the outdated (corresponding with the protagonist's) and the updated (corresponding with the participant's). Critically, these two perspectives compete for attention during processing (Altmann & Kamide, 2009; Rubio-Fernández, 2013). This attention-focus account is more parsimonious than the TB-default model defended by Leslie and colleagues.

Moreover, the attention-focus account can explain 3 year olds' success in FB tasks that allow children to focus their attention on the protagonist, without increasing the salience of the wrong response in the test phase (Rubio-Fernández & Geurts, 2013, 2015). As it was discussed in the Introduction, children's focus of attention in FB tasks can be an important predictor of their performance, yet it is not the only performance factor affecting young children (cf. Low, 2010; Peterson et al., 2005, 2012). However, young children's success in FB tasks that control the salience of the wrong response poses a challenge to Leslie's TB-default model, as well to the theory of FB reasoning proposed by Baillargeon, Scott, and colleagues (Baillargeon et al., 2010; Scott et al., 2010) insofar as they treat the TB response as a prepotent response and argue that successful performance in FB tasks requires an executive process of response inhibition (see de Bruin & Newen, 2012; Helming et al., 2014 for related critiques).

Supporting the traditional TB-bias view, a large number of studies have found a correlation between Theory of Mind and Executive Function (see Devine & Hughes, 2014 for a meta-analysis). The majority of these studies, however, used standard FB tasks to measure Theory of Mind development in children (see, e.g., Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Carlson, Mandell & Williams, 2004; cf. Perner et al., 2002). As it was argued earlier, standard FB tasks tend to increase the salience of the wrong response in the test phase (e.g., by mentioning the target object in the question). Most of the tasks used in these studies may have therefore artificially increased the need for inhibitory control in FB reasoning.

A more accurate assessment of the relationship between Theory of Mind and Executive Function would require using FB tasks that allow children to focus on the protagonist throughout the narrative, especially in the test phase (see Rubio-Fernández & Geurts, 2013, 2015). It is therefore an open empirical question whether children need to inhibit the TB response in order to pass an FB task that allows them to focus their attention on the protagonist throughout the story, without increasing the salience of the target object in the test phase.

10. Conclusions

When researchers started arguing in the 1990s that children under 4 years of age failed FB tasks because they attributed their own knowledge to the protagonist, all the available empirical evidence supported this claim (see Wellman et al., 2001). However, the experimental record of recent years calls for a revision of the traditional view: Numerous studies have now shown that infants are able to form the right expectations about the behavior of a character with a false belief (e.g., Senju, Southgate, Snape, Leonard, & Csibra, 2011; Southgate, Senju, & Csibra, 2007); 3 year olds have been shown to be able to pass a suitably streamlined version of the Sally–Anne task (Rubio-Fernández & Geurts, 2013, 2015), and the evidence from adult studies also runs counter to the TB-default model (Ferguson et al., 2010; Rubio-Fernández, 2013; Schneider, Bayliss, et al., 2012; Schneider, Lam, et al., 2012). Therefore, even if 3 year olds still fail the original version of the Sally–Anne task, the current experimental record challenges the traditional view and, more specifically, the TB-default model proposed by Leslie, Friedman, and colleagues.

Acknowledgments

This research was supported by a Marie Curie International Fellowship (Project 022149) and by a travel grant from the European Science Foundation (Experimental Pragmatics Programme). The author thanks Bart Geurts and Susanne Grassmann for many good discussions of the TB default and avoidance FB tasks.

References

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*, 55–71.

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adult performance on a non-inferential Theory of Mind task. *Cognition*, *106*, 1093–1108.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*, 841–844.

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, *115*, 54–70.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Science*, *14*, 110–118.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, *21*, 37–46.

Bartsch, K. (1996). Between desires and beliefs: Young children’s action predictions. *Child Development*, *67*, 1671–1685.

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*, 382–386.

de Bruin, L. C., & Newen, A. (2012). An association account of false-belief understanding. *Cognition*, *123*, 240–259.

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*, 1232–1254.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive Function and Theory of Mind: Stability and prediction from ages 2 to 3. *Developmental Psychology*, *40*, 1105–1122.

Carlson, S. M., & Moses, L. (2001). Individual differences in inhibitory control and children's Theory of Mind. *Child Development*, *72*, 1032–1053.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between Executive Function and Theory of Mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*, 73–92.

Coburn, P. I., Bernstein, D. M., & Begeer, S. (2015). A new paper and pencil task reveals adult false belief reasoning bias. *Psychological Research*, *79*, 739–749.

Deák, G. O., & Enright, B. (2006). Choose and choose again: Appearance-reality errors, pragmatics and logical ability. *Developmental Science*, *9*, 323–333.

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, *85*, 1777–1794.

Ferguson, H. J., Scheepers, C., & Sanford, A. J. (2010). Expectations in counterfactual and Theory of Mind reasoning. *Language and Cognitive Processes*, *25*, 297–346.

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*, 405–410.

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

Freeman, N. H., & Lacohée, H. (1995). Making explicit 3-year olds' implicit competence with their own false beliefs. *Cognition*, *56*, 31–60.

Friedman, O., & Leslie, A. M. (2004a). A developmental shift in processes underlying successful belief-desire reasoning. *Cognitive Science*, *28*, 963–977.

Friedman, O., & Leslie, A. M. (2004b). Mechanisms of belief-desire reasoning: Inhibition and bias. *Psychological Science*, *15*, 547–552.

Friedman, O., & Leslie, A. M. (2005). Processing demands in belief-desire reasoning: Inhibition or general difficulty? *Developmental Science*, *8*, 218–225.

Goldman, A. (2006). *Simulating minds: The philosophy, psychology and neuroscience of mindreading*. New York: Oxford University Press.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*, 26–37.

Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, *18*, 167–170.

Hoover, M. A., & Richardson, D. C. (2008). When facts go down the rabbit hole: Contrasting features and objecthood as indexes to memory. *Cognition*, *108*, 533–542.

Johansson, R., & Johansson, M. (2013). Look here, eye movements play a functional role in memory retrieval. *Psychological Science*, *25*, 236–242.

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentence with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, *38*, 1033–1050.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind." *Trends in Cognitive Sciences*, *8*, 528–533.

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief–desire reasoning as a process of selection. *Cognitive Psychology*, *50*, 45–85.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science*, *1*, 247–253.

Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: Relations with complex syntactical mastery. *Child Development*, *81*, 597–615.

MacDonald, M. C., & Just, M. A. (1989). Changes in activation level with negation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 633–642.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and Theory of Mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*, 622–646.

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

Neill, W. T., & Westberry, R. L. (1987). Selective attention and the suppression of cognitive noise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 327–334.

Nichols, S., & Stich, S. (2003). *Mindreading. An integrated account of pretence, self-awareness, and understanding of other minds*. Oxford, England: Clarendon Press.

Nickerson, R. S. (2001). The projective way of knowing. *Current Directions in Psychological Science*, *10*, 168–172.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the autistic child's Theory of Mind: Knowledge, belief, and communication. *Child Development*, *60*, 689–700.

Perner, J., Lang, B., & Kloo, D. (2002). Theory of Mind and self-control: More than a common problem of inhibition. *Child Development*, *73*, 752–767.

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child Development*, *76*, 502–517.

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development*, *83*, 469–485.

Petrashek, A. R., & Friedman, O. (2011). The signature of inhibition in Theory of Mind: Children's predictions of behavior based on avoidance desire. *Psychonomic Bulletin & Review*, *18*, 199–203.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood squares: Looking at things that aren't there anymore. *Cognition*, *76*, 269–295.

Rubio-Fernández, P. (2013). Perspective tracking in progress: Do not disturb. *Cognition*, *129*, 264–272.

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

Rubio-Fernández, P., & Geurts, B. (2015). Don't mention the marble! The role of attentional processes in false-belief tasks. *Review of Philosophy and Psychology*, Online First.

Ryskin, R. A., & Brown-Schmidt, S. (2014). Do adults show a curse of knowledge in false-belief reasoning? A robust estimate of the true effect size. *PLoS ONE*, *9*, e92406.

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General*, *141*, 433.

Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science*, *23*, 842–847.

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about identity at 18 months. *Child Development*, *80*, 1172–1196.

Scott, R. M., Baillargeon, R., Song, H. J., & Leslie, A. M. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, *61*, 366–395.

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science*, *22*, 878–880.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous Theory of Mind in Asperger syndrome. *Science*, *325*, 883–885.

Sommerville, J. A., Bernstein, D. M., & Meltzoff, A. N. (2013). Measuring beliefs in centimeters: Private knowledge biases preschoolers' and adults' representation of others' beliefs. *Child Development*, *84*, 1846–1854.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*, 587–592.

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.

Wellman, H. M., & Liu, D. (2004). Scaling of Theory of Mind tasks. *Child Development*, *75*, 523–541.

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