Journal of Experimental Child Psychology 177 (2019) 141–151

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Let's do it together! The role of interaction in false belief understanding

Marta Białecka-Pikul *, Magdalena Kosno, Arkadiusz Białek, Marta Szpak

Institute of Psychology, Jagiellonian University, 30-060 Kraków, Poland

ARTICLE INFO

Article history:
Received 3 January 2018
Revised 30 July 2018
Available online 8 September 2018

Keywords:
Theory of mind
False belief task
Interpersonal interaction

ABSTRACT

From the interaction theory perspective, we aimed to verify whether the interactive context in which theory of mind is measured can allow children younger than 4 years to pass the verbal false belief task (FBT). Therefore, an interactive FBT (iFBT) was devised, in which children were actively engaged in the story, and was used to test 210 children twice: first when they were 3 years old and again when they were 3.5 years old. Most 3-year-olds were unable to pass the iFBT (28% passed), but a half year later their performance was enhanced and they passed at a rate above chance level (59% passed). In addition, among 3.5-year-olds, passing the iFBT increases the chance of passing the standard FBT by four times. We conclude that the interactive mode of the FBT facilitates false belief understanding in children under 4 years of age.

Introduction

False belief understanding has been considered a "litmus test" for theory of mind (ToM) (for reviews, see Hughes & Devine, 2015; Rakoczy, 2017), and psychologists often use the so-called false belief task (FBT; Wimmer & Perner, 1983) to assess whether children have the ability to infer the mental states of others in order to predict their behavior. In the standard FBT, a character places a toy inside one container and goes out to play. In the meantime, a second character moves the toy to a different container. When the first character returns to the scene, children are asked to predict where the character will look for the toy. It is commonly believed that before their fourth birthday children are not able to pass this

* Corresponding author.
E-mail address: marta.bialecka-pikul@uj.edu.pl (M. Białecka-Pikul).

https://doi.org/10.1016/j.jecp.2018.07.018

task because they do not understand the first character's false belief (Wellman, Cross, & Watson, 2001). Since a prominent study appeared nearly 25 years ago (Clements & Perner, 1994), this view has been challenged in two ways. First, researchers have proposed that when nonverbal, less complex, and less cognitively demanding versions of the FBT are used (e.g., the violation of expectation paradigm), even toddlers and infants present spontaneous or implicit false belief understanding (see Buttelmann, Carpenter, & Tomasello, 2009; Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007; for reviews, see Heyes, 2014; Low & Perner, 2012). Second, Rubio-Fernández and Geurts (2013, 2016) argued that it is the attentional mechanism that is responsible for 3-year-olds' consistent failure on the standard verbal FBT. Here we propose a third alternative. We take a position based on interaction theory (for reviews, see Froese & Gallagher, 2012; Fuchs, 2013) and refer to pragmatic approaches in ToM research (Helming, Strickland, & Jacob, 2016; Westra & Carruthers, 2017) in explaining the important role of interaction in children's performance on the FBT. Moreover, a developmental perspective is taken, according to which it is argued that there exists a continuity in ToM development (Sodian, 2016; Thoermer, Sodian, Vuori, Perst, & Kristen, 2012); the development of basic communicative abilities in toddlers eventually allows older children to track the perspectives of others (Carpendale & Lewis, 2006; Liszkowski, 2013) during the FBT. Therefore, we postulate that when children actively participate in the FBT they can pass the task before their fourth birthday. Thus, the aim of the current research was to verify whether the interactive mode facilitates children's performance on the FBT.

Recently, Rubio-Fernández and Geurts (2013, 2016) provided strong evidence for the role of children's attentional processes during the FBT; allowing children to follow the character's behavior facilitates their performance on the FBT. Rubio-Fernández and Geurts (2013) proposed two important modifications of the FBT. The first modification involves the number of characters in the task and the way in which the displacement phase of the task is acted out. In the modified version of the FBT, called the Duplo task, a single character—a Duplo girl figure (i.e., made of large Lego blocks)—is present on the table throughout the telling of the story and, thus, could be observed by the child. The task begins when the Duplo girl puts the object in one of the boxes and then moves forward to the child and stays with her back to the boxes. The child is then asked about the character's perceptual access to the experimenter's activity, and if the child does not confirm the lack of the doll's access, the child is informed about it. At this point, the experimenter displaces the object to the second box behind the Duplo girl's back, and again a short exchange between the experimenter and the child takes place; the experimenter confirms that the character did not see the displacement. The second modification of the FBT is related to the test phase of this task. In the Duplo task, the child is not directly or indirectly asked about the false belief of the Duplo girl ("Where does she think the object is?" or "Where will she look for it?"); rather, the Duplo girl is handed to the child and the child is encouraged to show the experimenter: "What will happen next? What is she going to do now?"

In other words, two modifications to the standard FBT were applied. First, the story is about one character (instead of two characters) who is present while story is told. Second, the child is asked to act out the answer to an open-ended test question (instead of a standard question mentioning the object). Rubio-Fernández and Geurts (2013, 2016) showed that children are able to pass the Duplo task before their fourth birthday. According to the authors, these changes in the protocol allow children to follow the character's perspective and avoid priming the wrong answer, for example, by not mentioning the object in the test question (see Experiment 2b in Rubio-Fernández & Geurts, 2013, and both experiments in Rubio-Fernández & Geurts, 2016). We do not want to oppose this explanation, but we believe that an alternative explanation is also possible, especially concerning the function of the first modification in the FBT. We argue that the first modification in the Duplo task introduces an interactive or so-called "we-mode" (Gallotti & Frith, 2013; Tuomela, 2006) to the FBT. The child is no longer a passive observer of the story being presented by the experimenter but rather starts to be an interlocutor, a partner in a dialogue, who is asked and is informed—during the interaction—about the character's perceptual access ("Can the girl see me from where she is?" and "She hasn't seen what I did, has she?"). These short exchanges between the experimenter and the child about a third party (i.e., about the character's perceptual access), and the fact that in the context of these exchanges the experimenter displaces the object, may also *indirectly* inform the child about the experimenter's motive. In other words, the interactive mode or so-called we-mode is provided in the Duplo task. Children are actively engaged in the task; they actively participate rather than just observe the story being played out before them. We argue that such a reinterpretation of the Duplo task protocol is more widely adapted to interactive and pragmatic approaches to the study of ToM development.

According to interaction theory (Froese & Gallagher, 2012; Fuchs, 2013), our social understanding, or ToM, primarily occurs during interpersonal interactions. At least two lines of developmental research support this idea. First, young children's primary ability to be in tune with an adult, to share an interest, or to jointly attend to the same object and/or situation provides a basis for the development of social communication and cognition (Reddy, 2008; Trevarthen & Hubley, 1978). Second, in the domain of ToM research, the role of interaction in ToM development has been widely appreciated (e.g., Carpendale & Lewis, 2006). ToM in infants and toddlers develops through its interactive use (Liszkowski, 2013), or via the experience gained during communicative situations between a child and an adult, because the essential function of ToM is to predict the behaviors of others during social interactions (see Moll & Tomasello, 2007; Tomasello, Carpenter, Call, Behne, & Moll, 2005). Moreover, so-called pragmatic approaches to ToM (Helming et al., 2016; Westra & Carruthers, 2017) propose the same explanation, or a very similar one, for the inability of children younger than 4 years to pass the standard FBT. It is claimed that this failure is due to children's "failure to stick to a *third-person perspective* on the *instrumental* action performed by a mistaken agent while being requested to take a *second-person perspective* onto the experimenter's *communicative* action" (Helming et al., 2016, p. 449). We agree that the standard FBT is difficult for young children because these two contradictory demands are present: being in the we-mode with the experimenter (which is more natural to young children) while at the same time taking the third-person perspective (i.e., the character's perspective). Therefore, we modified the already interactive task proposed by Rubio-Fernández and Geurts (2013) to make it even more interactive and, thus, more suitable for younger children. We believe that this will allow us to observe enhanced performance on the FBT by children as young as 3.5 years even if the test question mentions the target object and directs children's attention to the wrong response.

How can one make the Duplo task more interactive? Two facts about the development of child–adult interaction are important here. First, amusement and humor are essential characteristics of early child–adult interactions (for a review, see Hoicka, 2014). Early peek-a-boo games with infants and more complex forms of teasing or pretending are natural parts of dialogues with infants and toddlers (Reddy, 2008). These dialogues are rewarding, and even 3-year-olds can understand nonliteral utterances such as jokes and ironic statements when they are provided in such a communicative context (Angeleri & Airenti, 2014). Second, deceptive practices are observed during interactions between adults and 2-year-olds (Newton, Reddy, & Bull, 2000). Moreover, children younger than 4 years can succeed in verbal false belief tasks when deception or trickery elements are introduced to the procedure (Chandler, Fritz, & Hala, 1989; Sodian, Taylor, Harris, & Perner, 1991; Sullivan & Winner, 1993). The deceptive motive of the experimenter also enhances children's performance on the FBT, especially when it is explicitly stated (see Wellman et al., 2001).[1] In other words, humor, provided by the trickery motive, is early developing and natural element of child–adult interaction, that can provide the child with a direct or indirect cue for understanding false belief with greater ease. By explicitly evoking a trickery motive or, more precisely, by evoking a *joint* trickery motive—that of the adult and the child together—that explains the experimenter's behavior in the interactive version of the FBT, we can make the empirical context of the FBT more conducive to success for young children (under 4 years), which will allow them to present their ToM ability.

Therefore, to further examine the role of interaction for children's successful performance on the FBT, the interactive FBT (iFBT) was prepared. Similar to the Duplo task, in our protocol there is one character only and the character does not leave the scene while the false belief narration is presented to the child. Most important, during the displacement phase of the iFBT the child is *directly* informed about the experimenter's intention to surprise the character (which in the iFBT is a puppet mouse). The experimenter emphasizes that "*we* are acting in secret" and says that "*we* will surprise the mouse" (for a detailed description of the procedure, see the "Measures" section in Method below). In other words, the experimenter evokes a joint trickery motive during the iFBT. These modifications were

[1] Rubio-Fernández and Geurts (2013) argued that in the Duplo task the "secret stance" is subtle and, thus, probably does not influence FBT performance. In our protocol, the secretive stance is directly stated and, thus, is made more salient.

intended to make our version more interactive than the Duplo task. However, because we are reconsidering the role of interaction during the FBT, and we do not want to focus on the role of the child's attention during the FBT as did Rubio-Fernández and Geurts (2016), we decided to retain the direct test question at the end of the task. We argue that if our modifications effectively facilitate the child's active engagement in interaction during the task, even if directly asking the child increases the salience of the wrong response (in line with the arguments of Rubio-Fernández & Geurts, 2016), the child will still be able to pass the task. In other words, although Rubio-Fernández and Geurts (2013, Experiment 2b) observed below-chance performance when asking the standard question, we expected that we would observe above-chance performance in 3.5-year-olds when the experimenter employs the explicit suggestion of trickery and speaks explicitly in the we-mode during the FBT (see the description of the experimenter's behaviors in the "Measures" section). If our hypothesis is correct, our results would show that active engagement with the experimenter can compensate for the distractive effect of mentioning the target object in the test question.

In addition, relative to the FBT, the iFBT contains one more modification to the test question. There were three reasons for this modification. First, pragmatic analyses of the standard FBT question (see Helming et al., 2016; Westra & Carruthers, 2017) pointed out that the standard question is *pragmatically infelicitous* for children younger than 4 years (i.e., the children may really be answering the question "Where should Sally look for the cookie in order to find it?" instead of "Where will Sally look for the cookie?"), and therefore it is necessary to be more precise when asking the question. Second, Siegal and Beattie (1991), as well as Yazdi, German, Defeyter, and Siegal (2006), showed that emphasizing the time point to which the test question refers—here, the starting point of the character's search—improves performance on the FBT. Third, because Rubio-Fernández and Geurts (2016) argued that the verb "to think" in the test question is more cognitively demanding than the verb "to look for" (p. 836), we decided to use behavior-related words when asking the test question (rather than verbs related to mental processes).[2] Therefore, instead of asking "where the character thinks the object is," we began by asking "where the character will *start* to *look for* the object" while stressing the word "start" with intonation and using it twice (see "Measures" section).

It is important to stress a few additional methodological points of the current study. In light of the current debate surrounding the "replication crisis"—with many researchers failing to replicate previous results in the area of implicit ToM (see Burnside, Ruel, Azar, & Poulin-Dubois, 2018; Crivello & Poulin-Dubois, 2018; Kulke, Reiß, Krist, & Rakoczy, 2018; Powell, Hobbs, Bardis, Carey, & Saxe, 2018)—our study could be considered as a conceptual[3] replication of Rubio-Fernández and Geurts (2013). The current study is not an exact replication given that our iFBT resembles the Duplo task only in regard to the first modification suggested by Rubio-Fernández and Geurts. However, by using a new task that is built on the design of the Duplo task, we further the discussion about the methodology of the verbal FBT and what aspects of task design are important for young children to succeed.

This is especially important because an exact replication of the study by Rubio-Fernández and Geurts (2013) partially failed (Kammermeier & Paulus, 2018); that is, the facilitating effect of the Duplo task as compared with the standard FBT was replicated, even though the above-chance performance in 3.5-year-olds, observed in the original study, was not present. Although the two studies by Kammermeier and Paulus included a larger number of children ($N = 79$ in both studies) than did the study by Rubio-Fernández and Geurts (2013) ($N = 22$), in both studies the ages of the participants varied considerably (i.e., by at least a few months; in Rubio-Fernández and Geurts the range was from 3.0 to 4.0 months, whereas in Kammermeier and Paulus the standard deviation was 3.6 months). Relative to these previous studies, we had the opportunity to test a large group of children ($N = 210$), to do so longitudinally,

---

[2] Although the meta-analysis by Wellman and colleagues (2001) found that neither temporal marking in the test question (e.g., "Where will Sally look *first*?") nor the exact wording of the test question (e.g., "Where will Sally look for the cookie?" vs. "Where does Sally think the cookie is?") has an effect on the results on the FBT. In light of the above three reasons, we decided to simplify the question by using "look for" instead of "think" and by marking the starting point.

[3] Schmidt (2009) clearly distinguished the "narrow bounded notion of direct/exact replication" (i.e., the "repetition of experimental procedures") from the wider notion of replication, which can be called conceptual when researchers repeat "a test of a hypothesis or a result of an earlier research work with different methods" (p. 91). Importantly, contrary to direct replication, which aims at the confirmation of facts and the extension of knowledge, the scientific gain of conceptual replication is deeper understanding of the underlying mechanism that is crucial for science.

and to provide a very narrow age range (i.e., our 3.5-year-olds are really that age, with a standard deviation of only 0.46 months). In this way, the current study may help to rule out the possibility that age differences are at least partly responsible for the results obtained in the previous studies.

In addition, we also followed Rubio-Fernández and Geurts (2013) by using the Smarties task—that is, the change of content FBT—rather than the standard change of location task. This is a very important methodological decision, especially in light of the argument that the more natural benchmark task would be the standard change of location task. There were three main reasons why we used the change of content task. First, we planned to compare our results with those of Rubio-Fernández and Geurts (2013) because we saw our study as a conceptual replication of their research, even if our main aim was to reinterpret the results in light of interaction theory. Second, Wellman et al. (2001) found that the type of task (change of location vs. change of content) has no effect on the results of the FBT. Third, the results obtained using the change of location FBT and the change of content FBT were reported to be highly correlated (Carlson & Moses, 2001; Grosse Wiesmann, Friederici, Singer, & Steinbeis, 2017).

Three main hypotheses were the subject of empirical verification in the current study. First, assuming that children's active engagement in the FBT helps children to pass the task, we expected that 3.5-year-olds would pass the iFBT at a higher rate than they pass the standard FBT, which here was the Smarties task. Second, because this interactive mode is natural even during the earliest developmental exchanges between children and adults, the iFBT is an easier empirical context for capturing ToM than the standard FBT. Moreover, because performance on the iFBT should be related to performance on the Smarties task given that these tasks are versions of the change of location and change of content FBTs that have been reported to correlate (Carlson & Moses, 2001; Grosse Wiesmann et al., 2017), we expected that performance on the iFBT would predict the performance of 3.5-year-olds on the Smarties task. Third, because there is a clear developmental change in FBT performance (Wellman et. al., 2001), it was also expected that the same developmental pattern would be observed for the iFBT and, thus, more 3.5-year-olds than 3-year-olds would pass the iFBT.

Method

Procedure

The study was part of a larger 3-year longitudinal research project [The birth and development of mentalizing abilities] that received clearance from the institutional ethics board. The children were recruited on a voluntary basis via personal advertisements. Informed consent was received from all caregivers. Children visited the laboratory six times (every 6 months). Participants first visited the lab when the children were 12 months old, and they visited for the last time when they were 42 months old. Children received a small gift at the end of each lab visit. Children were tested individually, and each laboratory session lasted approximately 1 h.

As regards the current study, only false belief measures obtained during two lab visits were analyzed. It should be noted that during the first lab visit we tested 361 1-year-olds, and during the last lab visit we tested 299 3.5-year-olds. The iFBT was used with the 3-year-olds (here, Time 1 [T1]). Six months later, at 3.5 years of age (Time 2 [T2]), the same children participated in a second lab visit when two FBTs were administered: the iFBT and the Smarties task. The FBTs were always administered among other tasks not reported here. Importantly, during T1 children participated in 26 different tasks in two orders, and the iFBT was administered as the 10th task in the first order and as the 14th task in the second order (the results did not differ between the two orders, $\chi^2 = 0.114$, $p = .73$). During the lab visit at 3.5 years of age, the session consisted of 23 tasks, and the iFBT was administered as the 8th task and the Smarties task as the 13th task (always in this order). Precise descriptions of all tasks that measured language and communicative abilities, executive functions, and theory of mind are not presented here because we do not refer to their results.

Sample

A total of 312 3-year-old children participated at T1 (M_{age} = 35.52 months, SD = 0.47). From this original sample, 269 3-year-olds were tested with the iFBT at T1 (M_{age} = 35.51 months, SD = 0.49).

The dropouts were due to a lack of cooperation, tiredness, or technical problems related to recording the laboratory visit. During T2, 219 children from T1 completed the iFBT. The attrition rate between T1 and T2 was 18.6%. Following the protocol in Rubio-Fernández and Geurts (2013), children who gave two responses (correct and incorrect) in the iFBT were excluded from further analysis. Thus, the final sample consisted of 210 children (91 girls, 43.3%) who completed the iFBT measures at both time points (T1: M_{age} = 35.50 months, *SD* = 0.46, range = 34.52–39.05; T2: M_{age} = 41.49 months, *SD* = 0.46, range = 40.03–44.01). The results obtained from the children who were excluded from the analysis at T1 (n = 59) did not differ from the results obtained from the children who were included in the analysis, χ^2 = 1.945, p = .16. The majority of the children were from urban areas within Poland (73.7%), and more than half of the parents were educated to a university degree level (56.3%).

Measures

Interactive false belief task

The iFBT was based on the procedure proposed by Rubio-Fernández and Geurts (2013). The child sat at a table (50 cm high × 188 cm wide × 74 cm deep), the parent was behind the child, and the experimenter was in front of the child. There were two boxes (red and blue), one on the left side of the table and the other on the right side (90 cm apart). The boxes were 35 cm from the experimenter and out of the child's reach. Each of the boxes (20 cm long × 10 cm wide × 10 cm high) contained a small video (dashboard) camera that recorded the child's gazes and gestures. The whole scene was recorded by two video (CCTV) cameras placed in the corners of the room. The parents were instructed to remain silent and neutral throughout the session. The following props were used in the task: a puppet mouse, a plastic cookie, and a small blanket.

The task started with the introduction of the mouse. The experimenter put a puppet on her hand, showed that the mouse was holding the cookie, and said, "Oh look, it is the mouse. Listen to this—I will tell you about this mouse." When the child was concentrating on the puppet, the story was continued: "The mouse likes cookies very much. Today the mouse has eaten some of the cookie (the experimenter makes the puppet eat), and now the mouse is putting the cookie in the box." At this point, the experimenter made the mouse put the cookie inside one of the two boxes (the choice of the box was counterbalanced across participants). The experimenter commented on this in the following way: "You see, the mouse put the cookie in the box," and then moved the mouse toward the child and said, "Now the mouse is going to sleep." At this point, the experimenter put the mouse on the table (close to the child), covered the mouse with a blanket, and said, "Look, the mouse is sleeping." Then the child was asked, "Can the mouse see me?" This question was only a prompt; thus, regardless of the child's answer, the experimenter said, "No, the mouse can definitely not see or hear me. The mouse is sleeping deeply, the mouse will not see what I do." The experimenter then added with a secretive tone in her voice, "Hey, let's surprise the mouse!" This direct appeal was accompanied by nonverbal cues. The experimenter was acting "in secret," or suggesting connivance, by speaking in a subdued voice and by crouching and leaning forward toward the child when saying "Let's surprise the mouse!" The experimenter then moved the cookie from one box to the other. At this point, the experimenter showed the mouse and said, "She didn't see what I did, did she?" This too was only a prompt. Regardless of the child's reaction, the experimenter said, "No, she definitely didn't see what happened because she is sleeping!" Once it had been established that the mouse had not seen the replacement, the experimenter made the mouse wake up and commented, "Oh, the mouse has woken up!" After putting the blanket away and putting the puppet on her hand, the experimenter turned the mouse toward the boxes and said, "Now the mouse *is starting to* look for her cookie. She has to look for it; where does she *start* to look?" If the child did not answer within 5 s, the experimenter encouraged the child, saying, "Show me." As we mentioned previously, there were a handful of important similarities and differences between the current task and the Duplo task (Rubio-Fernández & Geurts, 2013, Experiments 1, 2a, and 2b). First, as in the Duplo task, we also had only one character visible to the child throughout the narration, not two characters. Second, we also made the child focus on the mouse during the transfer of the cookie by checking that the mouse could not see what was going on (see Rubio-Fernández &

Geurts, 2013, Experiment 2a). At this point in the task, we also highlighted that the mouse was sleeping and could not see or hear anything under the blanket. Third, unlike in the original Duplo task, we asked a direct test question and did not ask the child to act out his or her responses with the mouse (see Rubio-Fernández & Geurts, 2013, Experiment 2b). The reason for using a direct test question is that even if it mentioned the object, we expected that it would not be detrimental to the child's responses because our interactive modification was expected to exceed the distracting effect of the test question. Moreover, we could compare our results with those from the Smarties task as in Rubio-Fernández and Geurts (2013). Fourth, unlike in previous studies with the Duplo task, the experimenter explicitly stressed the "secret stance" by directly saying that the character would be tricked by the experimenter and the child together, something that was highlighted by the use of additional nonverbal cues (i.e., tone of voice and posture). Critically, we predicted that this last manipulation should make the task more interactive, which should in turn help the child to be more engaged in the deception than in the version of the Duplo task that used a direct test question (Rubio-Fernández & Geurts, 2013, Experiment 2b).

The child's first answers on the iFBT were coded as either a gesture or a verbal answer. The child was credited with 1 point when he or she pointed to or named the correct box. If the child pointed to or named the wrong box, 0 points were assigned. All ambiguous answers (e.g., when a child pointed to the correct box but then changed his or her mind and pointed to the wrong box) were excluded (see "Procedure" section above). Two independent coders scored 20% of materials and yielded satisfactory reliability at both T1 and T2 measurement points, kappas = .79 and .88, respectively.

The Smarties task

In the Smarties task (called the contents false belief task; Wellman & Liu, 2004), the child was shown a prototypical box that usually contains toothpaste. The experimenter was sitting opposite the child, and the toothpaste box, a toothbrush, and a cup were on the table. The experimenter started by showing the child a puppet named Bert, saying that Bert would join them in a minute but that Bert was now in the cupboard and could not hear or see what was happening on the table. The experimenter put Bert in the cupboard behind her back. The experimenter then asked the child what is usually found in that kind of a box (indicating the toothpaste box). After the child had answered, the box was opened and the child was shown the true content—a balloon. The experimenter then put the balloon back in the box, closed the box carefully, and asked the child a question: "What is inside this box?" Next, the experimenter took Bert out of the cupboard and said, "Now Bert is coming out of the cupboard, and he has not looked inside this box before. So, what does Bert think is inside the box?" (the target question). If the child did not answer after 3 s, the experimenter asked the prompt question: "Does Bert think there's toothpaste inside or a balloon?" After the child had reacted, a memory question was also asked: "Has Bert looked inside this box before?"

The child was credited with 1 point if he or she answered the target or prompt question correctly and—as in the standard version of this task—did not fail the memory question. To make the scoring of the Smarties task comparable to the scoring of the iFBT, we did not take into account the answers to the memory question and excluded ambiguous answers (e.g., no reactions).[4] However, if the child gave a correct answer to the target or prompt question, 1 point was awarded, and if the answer was incorrect, 0 points were awarded. Two independent coders scored 20% of materials and yielded perfect reliability, kappa = 1.00.

Results

Regarding our first hypothesis, we compared the performance of 3.5-year-olds on the iFBT and the Smarties task. Among children who took part in the iFBT (N = 210), 194 also took part in the Smarties

[4] We are grateful to an anonymous reviewer for this suggestion. We excluded the results of 16 children in the Smarties task who gave ambiguous answers. The group we tested for this task was N = 194. Moreover, we checked that the results following the exclusion of these children did not differ from the results of the rest of the group on the iFBT, ^{2} = 0.11, p = .740.

task. As expected, the Smarties task was more difficult for children than the iFBT; whereas 73 3.5-year-olds passed the Smarties task (38% success rate), 123 passed the iFBT (59% success rate). A McNemar test with continuity correction revealed a significant difference in participants' performance on the iFBT and on the Smarties task, $\chi^2 = 18.93$, $p < .001$. Children performed significantly below chance level on the Smarties task ($p < .001$, two-choice binomial test, two-tailed) and significantly above chance level on the iFBT ($p = .016$, two-choice binomial test, two-tailed).

Furthermore, we used a logistic regression analysis to test our second hypothesis and verify whether performance on the iFBT was a predictor of performance in the Smarties task. The model for the Smarties task was statistically significant, $\chi^2 = 19.79$, $p < .001$. The iFBT at T2 was a significant predictor of children's performance in the Smarties task at T2, Negelkerke's $R^2 = .13$, $B = 1.42$, $Wald = 17.74$, $p < .001$. The Exp(B) value indicated that passing the iFBT increased the chance of passing the Smarties task by four times, Exp(B) = 4.135.

Moreover, in line with our third hypothesis, older children were better in false belief understanding than younger children; at T1 58 children passed the task (27.6% success rate), whereas at T2 123 children passed the task (58.6% success rate). A McNemar test with continuity correction revealed a significant difference in children's performance, $\chi^2 = 39.77$, $p < .001$. Children who were 3 years old performed significantly below chance level on the iFBT ($p < .001$, two-choice binomial test, two-tailed), whereas 3.5-year-olds performed significantly above chance level (see above).

In addition, our analysis of children's performance on the iFBT at the two measurement points revealed that 107 children had the same results at T1 and T2; that is, 68 children did not pass the task at either time point, and 39 passed it at both time points. Among 84 children, a positive change in iFBT performance was observed (from not passing the task to passing the task), whereas 19 children showed a decrease in their iFBT performance. The sign test indicated that positive changes were higher than negative changes, $Z = -6.31$, $p < .001$. Thus, this shows that there was a general positive age-related change in iFBT performance; most of the children who gave incorrect answers at T1 were able to pass the task at T2.

Discussion

There is a common belief that children younger than 4 years are not able to pass the standard verbal FBT. However, it has also been shown that subtle changes in the FBT protocol can influence children's performance (Rubio-Fernández & Geurts, 2013, 2016; Wellman et al., 2001). Taking the stance of interaction theory and pragmatic approaches in ToM research, the aim of the current study was to check whether the interactive mode of the FBT would allow children younger than 4 years to show their ToM abilities.

By use of the iFBT, we succeeded in extending the results obtained by Rubio-Fernández and Geurts (2013, 2016) with a group of more than 200 children. Therefore, we have provided more data showing that active participation in the FBT allows children to pass the task before their fourth birthday. In the current study, we tested a large group of children in a narrow age range, thereby eliminating possible age-related variation in FBT performance. Thus, we can confidently assert that roughly 60% of children at 3.5 years of age are able to pass the iFBT. With a comparable protocol, we conceptually replicated the results of Rubio-Fernández and Geurts (2013, 2016). Our results, show that 3.5-year-olds are able to respond correctly to a standard test question if the FBT is sufficiently interactive.

Because our main aim was a reinterpretation and conceptual replication of Rubio-Fernández and Geurts's (2013, 2016) studies, it can be concluded that the current results support the idea that joint child–adult interaction (see Carpendale & Lewis, 2006) plays an important role in ToM development. From a methodological point of view, we also found that modifying the FBT to make it more ecologically valid, more child friendly, and more interactive allows children younger than 4 years to pass it even if they are asked a direct test question. This extends the original results by Rubio-Fernández and Geurts (2013, 2016) by showing that children's engagement in the task can help them to overcome distractions associated with the mention of the target object (what Helming et al. (2016) referred to as a "referential bias").

In the standard unexpected transfer task, the child is a passive observer of the testing situation. By contrast, within the iFBT, the child is an active participant in dialogue with an adult who has a trickery motive, namely, to surprise the character. Moreover, this motive was presented as a joint motive shared by the child and the experimenter: "Hey, let's surprise the mouse!" This modification of the standard FBT transformed the testing situation into a more natural interaction during which the child is actively engaged (Airenti, 2015). Indeed, nearly 30% of 3-year-olds and nearly 60% of 3.5-year-olds passed the iFBT, whereas about 40% of 3.5-year-olds passed the FBT (i.e., Smarties task).

Our results showed that at 3.5 years of age children's performance on the iFBT was a significant predictor of their performance on the Smarties task. Passing the iFBT increases the chances of passing the Smarties task by four times. As was assumed, the iFBT was an easier empirical method for capturing young children's developing ToM. What is more, by using the iFBT we showed that older children (3.5-year-olds) were better in false belief understanding than younger children (3-year-olds). Older children performed significantly above chance level on the iFBT and, importantly, most children who failed the iFBT at 3 years of age were able to pass the task at 3.5 years of age.

The results of this study are in line with an interactive and social constructivist account of ToM development (Carpendale & Lewis, 2006). The abilities to act together and to form and understand joint intentions are considered to be necessary preconditions for coming to understand false beliefs (Liszkowski, 2013; Moll & Tomasello, 2007; Tomasello et al., 2005). By making use of only one character during the iFBT, directly inviting the child to participate in the interpersonal interaction, and evoking pretense and establishing the social context of joint trickery, a feeling of togetherness is created. Moreover, these modifications lead to transforming the testing situation from a third-person perspective to a second-person perspective (Fuchs, 2013). We found that such an interactive mode facilitates FBT performance. In addition, on the one hand, we confirmed previous observations that trickery or deceptive motivations increase FBT performance (Sullivan & Winner, 1993); on the other hand, we devised a more natural and developmentally valid version of the FBT—one that is more suitable for children younger than 4 years and does not contain the demands of taking both third-person and second-person perspectives at the same time (Helming et al., 2016).

The current study has a design-related limitation that should be addressed in future research. Although the standard FBT was included, the nonexperimental design of the current study did not allow for the direct testing of the impact of all procedural factors that may have played a role in children's ability to pass the FBT. First of all, it would be important to more directly compare our results and interpretations with those obtained and made by Rubio-Fernández and Geurts (2013, 2016). According to Rubio-Fernández and Geurts, the direct question found during the test phase of the FBT makes the task more difficult because mentioning the target object primes the wrong answer. Contrary to this reasoning, in our study the presence of the direct question naming the target object at the end of the iFBT did not seem to interfere with children's performance given that nearly 60% of 3.5-year-olds were able to pass the task. It is worth mentioning that in a study including an open-ended question in the Duplo task (Rubio-Fernández & Geurts, 2013, Experiment 1), nearly 80% of the 3.5-year-old participants passed, but in a study with the standard question in the Duplo task (Rubio-Fernández & Geurts, 2013, Experiment 2b), only 22% passed. There are probably two reasons why our results fall in between these findings. First and foremost, in accordance with interaction theory (e.g., Froese & Gallagher, 2012), we helped the children to pass the task by making it more interactive (via directly stating the joint trickery motive and the "we-mode" of interaction). Second, our way of mentioning the target object in the test question might actually not be so detrimental for the task performance. Asking "where will the mouse *start to* look for . . ." was probably more pragmatically suitable for younger children (see Yazdi et al., 2006). This interpretation should be verified in further experimental research by manipulating only one factor and leaving the others constant. For example, we may conduct the iFBT with an open-ended question, not mentioning the target object as in the Duplo task. Alternatively, we could conduct the Duplo task with our easier "start to look for" question.

In addition, it should be emphasized that both our iFBT and the Duplo task differ from the Smarties task in many dimensions (e.g., content vs. location, explicit deception vs. no deception). Therefore, one may claim that the difference in children's performance in those tasks is related to other factors than those that we assume are making the tasks interactional.[5] Thus, it would be prudent to devise two versions of the FBT that would be the same on all dimensions but differ only in their interactional factors. This would allow for the comparison of children's performance on an interactional FBT and a noninteractional FBT. We are convinced that such direct comparisons would further confirm the thesis that, from a pragmatic or interactive point of view, there are various ways in which to make even the standard verbal FBT more suitable for young children.

Conclusions

Despite these limitations, this study provides important data. First, these longitudinal data, gathered from a large group of children, shows a developmental increase in performance on the interactive FBT. We argue that further support for the importance of social interaction in the development of false belief understanding is found in the fact that performance on the iFBT is predictive of performance on the standard FBT. We may claim, therefore, that "doing the task together" is an important factor influencing the performance of children younger than 4 years on the false belief task.

Acknowledgments

This research was supported by grants from the Polish National Science Centre: "The Birth and Development of Mentalising Ability" (2011/01/B/HS6/00453) and "Stability and Continuity in the Development of Theory of Mind in Middle Childhood: Trajectories and Predictors of Development" (2015/19/B/HS6/01252). We express our gratitude to all the children and parents who participated in the study. We also thank all the team members for their hard work in collecting and coding the data. We are especially grateful for the important comments and suggestions on the first draft of the manuscript provided by Paula Rubio-Fernández.

References

Airenti, G. (2015). Theory of mind: A new perspective on the puzzle of belief ascription. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01184.
Angeleri, R., & Airenti, G. (2014). The development of joke and irony understanding: A study with 3- to 6-year-old children. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 68*, 133–146.
Burnside, K., Ruel, A., Azar, N., & Poulin-Dubois, D. (2018). Implicit false belief across the lifespan: Non-replication of an anticipatory looking task. *Cognitive Development, 46*, 4–11.
Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112*, 337–342.
Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.
Carpendale, J., & Lewis, C. (2006). *How children develop social understanding*. Malden, MA: Blackwell.
Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development, 60*, 1263–1277.
Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377–395.
Crivello, C., & Poulin-Dubois, D. (2018). Infants' false belief understanding: A non-replication of the helping task. *Cognitive Development, 46*, 51–57.
Froese, T., & Gallagher, S. (2012). Getting interaction theory (IT) together: Integrating developmental, phenomenological, enactive, and dynamical approaches to social interaction. *Interaction Studies, 13*, 436–468.
Fuchs, T. (2013). The phenomenology and development of social perspectives. *Phenomenology and the Cognitive Sciences, 12*, 655–683.
Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences, 17*, 160–165.
Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science*. https://doi.org/10.1111/desc.12445.
Helming, K. A., Strickland, B., & Jacob, P. (2016). Solving the puzzle about early belief-ascription. *Mind & Language, 31*, 438–469.
Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science, 17*, 647–659.
Hoicka, E. (2014). The pragmatic development of humor. In D. Matthews (Ed.). *Pragmatic development in first language acquisition* (Vol. 10, pp. 219–237). Amsterdam: John Benjamins.

[5] We are grateful to two anonymous reviewers for this suggestion.

Hughes, C., & Devine, R. T. (2015). A social perspective on theory of mind. In M. Lamb & R. M. Lerner (Eds.), Handbook of child psychology and developmental science (7th ed.), Vol. 3: Social, emotional and personality development. Hoboken, NJ: John Wiley.
Kammermeier, M., & Paulus, M. (2018). Do action-based tasks evidence false-belief understanding in young children? *Cognitive Development, 46*, 31–39.
Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of theory of mind? Replication attempts across the life span. *Cognitive Development, 46*, 97–111.
Liszkowski, U. (2013). Using theory of mind. *Child Development Perspectives, 7*, 104–109.
Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology, 30*, 1–13.
Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 362*, 639–648.
Newton, P., Reddy, V., & Bull, R. (2000). Children's everyday deception and performance on false-belief tasks. *British Journal of Developmental Psychology, 18*, 297–317.
Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258.
Powell, L. J., Hobbs, K., Bardis, A., Carey, S., & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development, 46*, 40–50.
Rakoczy, H. (2017). Theory of mind. In B. Hopkins, E. Geangu, & S. Linkenauger (Eds.), *Cambridge encyclopedia of child development* (pp. 505–514). Cambridge, UK: Cambridge University Press.
Reddy, V. (2008). *How infants know minds*. Cambridge, MA: Harvard University Press.
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. (2016). Don't mention the marble! The role of attentional processes in false-belief tasks. *Review of Philosophy and Psychology, 7*, 835–850.
Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*, 90–100.
Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition, 38*, 1–12.
Sodian, B. (2016). Is false belief understanding continuous from infancy to preschool age? In D. Barner & A. S. Baron (Eds.), *Core knowledge and conceptual change*. Oxford, UK: Oxford University Press.
Sodian, B., Taylor, C., Harris, P. L., & Perner, J. (1991). Early deception and the child's theory of mind: False trails and genuine markers. *Child Development, 62*, 468–483.
Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*, 587–592.
Sullivan, K., & Winner, E. (1993). Three-year-olds' understanding of mental states: The influence of trickery. *Journal of Experimental Child Psychology, 56*, 135–148.
Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology, 30*, 172–187.
Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). In search of the uniquely human. *Behavioral and Brain Sciences, 28*, 721–727.
Trevarthen, C. H., & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding, and acts of meaning in the first year. In A. Lock (Ed.), *Action, gesture, and symbol: The emergence of language*. London: Academic Press.
Tuomela, R. (2006). Joint intention, we-mode, and I-mode. *Midwest Studies in Philosophy, 30*(1), 35–58.
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.
Westra, E., & Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition, 158*, 165–176.
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.
Yazdi, A. A., German, T. P., Defeyter, M. A., & Siegal, M. (2006). Competence and performance in belief–desire reasoning across two cultures: The truth, the whole truth, and nothing but the truth about false belief? *Cognition, 100*, 343–368.