Journal of Experimental Child Psychology 177 (2019) 297–312

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Memory and inferential processes in false-belief tasks: An investigation of the unexpected-contents paradigm

Paula Rubio-Fernández *

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA
Department of Philosophy, University of Oslo, Blindern, NO-0315 Oslo, Norway

A R T I C L E I N F O

Article history:
Received 26 January 2018
Revised 19 August 2018
Available online 27 September 2018

Keywords:
False-belief task
Unexpected-contents task
Associative memory
Belief inference
Two-system account of Theory of Mind
Replication studies

A B S T R A C T

This study investigated the extent to which 3- and 4-year-old children may rely on associative memory representations to pass an unexpected-contents false-belief task. In Experiment 1, 4-year-olds performed at chance in both a standard Smarties task and a modified version highlighting the secrecy of the contents of the tube. These results were interpreted as evidence that having to infer the answer to a false-belief question (without relying on memory representations) is generally difficult for preschool children. In Experiments 2a, 2b, and 2c, 3-year-olds were tested at 3-month intervals during their first year of preschool and showed better performance in a narrative version of the Smarties task (chance level) than in the standard version (below-chance level). These children performed even better in an associative version of the narrative task (above-chance level) where they could form a memory representation associating the protagonist with the expected contents of a box. The results of a true-belief control suggest that some of these children may have relied on their memory of the protagonist's preference for the original contents of the box (rather than their understanding of what the protagonist was expecting to find inside). This suggests that when 3-year-olds passed the associative unexpected-contents task, some may have been keeping track of the protagonist's initial preference and not only (or not necessarily) of the protagonist's false belief. These results are interpreted in the

* Address: Department of Philosophy, University of Oslo, Blindern, NO-0315 Oslo, Norway.
E-mail address: paula.rubio-fernandez@ifikk.uio.no

https://doi.org/10.1016/j.jecp.2018.08.011

light of current accounts of Theory of Mind development and failed replications of verbal false-belief tasks.

Introduction

Since researchers discovered that infants were able to pass suitably adapted false-belief tasks, the main debate in Theory of Mind has concerned how babies are able to pass such tasks; that is, whether infants are indeed able to attribute false beliefs to others (e.g., Baillargeon, Scott, & He, 2010; Carruthers, 2013; Helming, Strickland, & Jacob, 2014) or they pass false-belief tasks by relying on lower-level abilities (e.g., Apperly & Butterfill, 2009; Low & Perner, 2012; Ruffman, 2014). This question, however, is relevant not only to the interpretation of infant studies but also to young children's performance on verbal false-belief tasks. The current study investigated whether 3- and 4-year-old children may rely on associative memory processes to pass such tasks.

The Sally–Anne task is the archetypal change-of-location false-belief task (Baron-Cohen, Leslie, & Frith, 1985). In this task, a puppet named Sally puts a marble in a box before going out to play. In her absence, a puppet named Anne moves the marble to a basket. The child is then asked the standard test question: "When Sally comes back, where will she look for her marble?" Hundreds of Theory of Mind studies during the past 35 years have shown that children are not able to pass change-of-location tasks before their fourth birthday, with younger children predicting that Sally will look for her marble in its current location rather than in the container where she left it (Wellman, Cross, & Watson, 2001).

Another standard false-belief task is the unexpected-contents task. In the Smarties task (Perner, Frith, Leslie, & Leekam, 1989), for example, children are shown a tube of Smarties (a familiar box of chocolates) whose actual contents is unconventional (e.g., pencils), and they need to predict what somebody else would say is inside the tube (e.g., "When the next kid comes in and we ask him what's in here, what is he going to say?"). As in change-of-location tasks, children under 4 years of age tend to fail unexpected-contents tasks, predicting that the next kid will say that there are pencils in the tube (as if the kid were knowledgeable).

Because 3-year-old children perform comparably in change-of-location and unexpected-contents tasks (Wellman et al., 2001), these two types of false-belief task are often treated as equivalent tests and have often been combined in Theory of Mind studies for a more robust measure of belief understanding. However, although both tasks were designed to investigate belief attribution, these two tasks pose very different demands, potentially affecting young children's performance differentially.

One difference previously discussed in the literature is that the two possible responses to the false-belief question are not equally salient in change-of-location and unexpected-contents tasks (Freeman & Lacohée, 1995; Mitchell & Lacohée, 1991; Rubio-Fernández & Geurts, 2016). Whereas in the Sally–Anne task children must choose between two locations in the scene when predicting where Sally will return for her marble, in the Smarties task, only the wrong response is physically present in the setting; that is, there are no Smarties in the scene, only pencils. Early studies have shown that physically representing the two possible responses to the false-belief question in an unexpected-contents task improves performance in younger groups, probably because it reduces the salience of the wrong response (Freeman & Lacohée, 1995; Mitchell & Lacohée, 1991).

Another difference between change-of-location and unexpected-contents tasks is that in the former, but not in the latter, children must follow a series of events in a narrative and adopt the protagonist's perspective during the test phase. By contrast, in unexpected-contents tasks, children need to predict what another child would say when the child is asked the test question, but this other child was not previously introduced to the participant and did not play any role in the task up to the test phase. Thus, the design of change-of-location tasks should in principle allow children to track a protagonist's perspective throughout the narrative, whereas this would not be possible in unexpected-contents tasks.

Lewis, Freeman, Hagestadt, and Douglas (1994; see also Lewis, 1994) reported a series of studies showing how a narrative version of the standard change-of-location task elicited improved performance in children under 4 years of age by allowing them to focus on the causal actions that built the false-belief story (rather than the current location of the object). According to Lewis and colleagues, 3-year-olds tend to fail false-belief tasks because they are unable to retrace vital links between the events and the mental states of the protagonist, thereby failing to derive the necessary inferences when responding to the false-belief question. Narrative tasks that strengthen such links between the story events and the protagonist's perspective should be more accessible for younger children.

Rubio-Fernández and Geurts (2016) pointed out that the narrative format of change-of-location tasks would allow young children to form a memory trace associating the protagonist with the object in its original location. It is important to note that this memory trace is necessary to pass a change-of-location task because it precedes the attribution of a false belief to the protagonist when the object changes location. However, this associative memory representation might also allow young children to correctly predict the protagonist's return without necessarily attributing a false belief to the protagonist (an interpretation also considered by Lewis, 1994, p. 469). Because the format of unexpected-contents tasks does not allow children to form an association between the agent and the correct response to the test question, Rubio-Fernández and Geurts (2016) argued that unexpected-contents tasks should be harder than change-of-location tasks — especially if young children rely on associative memory representations to pass modified false-belief tasks. The current study aimed to test Rubio-Fernández and Geurts's prediction.

Two modified false-belief tasks with rich and lean interpretations

The Duplo task is a variation on the classic Sally–Anne task that introduces two sets of modifications to the original paradigm, both intended to help the child stay focused on the protagonist's perspective (Rubio-Fernández & Geurts, 2013, 2016). First, it is ensured that the child can see the protagonist throughout the displacement phase of the task. Rather than making the figure disappear, as is standardly done in this type of false-belief task, the experimenter makes the Duplo girl walk in the direction of the child and turn her back on the scene. As the experimenter transfers the girl's bananas from one container to the other, the experimenter keeps checking with the child that the protagonist cannot see what the experimenter is doing.

The second set of modifications is introduced during the test phase. When the experimenter returns the figure back to the center of the scene, rather than asking the child the standard false-belief question, the experimenter places the figure in front of the two cupboards and asks whether the child would like to continue the story. The experimenter then encourages the child to take the lead by asking, "What happens next? What is the girl going to do now?"

Three-year-old children were able to pass the Duplo task (Rubio-Fernández & Geurts, 2013, 2016), performing significantly above chance rather than below (as they normally do at that age). The results of these two studies further showed that both sets of modifications are crucial for 3-year-olds' success in the Duplo task; that is, young children need to be continuously focused on the protagonist in order to predict that she will return to the empty container. Making the girl figure disappear before the transfer of the object (as is normally done in the Sally–Anne task) is sufficient for young children to fail the Duplo task. Likewise, drawing children's attention to the bananas before they need to continue the story (e.g., by asking children, "Where are the bananas now?" or "Now the Duplo girl is hungry and wants a banana"), or simply mentioning the bananas in the standard test question (e.g., "Where will the Duplo girl look for her bananas?"), also disrupts young children's focus on the protagonist and leads them to predict the wrong outcome.

The design of the Duplo task is in line with Lewis et al.'s (1994) narrative version of a standard change-of-location task in that it allows children to follow the false-belief plot without losing track of the protagonist's perspective.[1] However, like the results of Lewis et al.'s study, 3-year-olds' successful

[1] I thank an anonymous reviewer for pointing out the parallels between Lewis et al.'s (1994) early work on narrativity and false-belief tasks and the aims and design of the Duplo task.

performance in the Duplo task can be explained more or less conservatively. On a rich reading, the design of the Duplo task would allow children under 4 years of age to reveal their true understanding of beliefs (previously masked by the inappropriate design of standard tasks). On a less assuming reading, the design of the Duplo task would allow children to focus on the protagonist and the location where she left the target object without necessarily attributing a false belief to the protagonist when correctly predicting her return.

A recent study by Buttelmann, Over, Carpenter, and Tomasello (2014) used a modified unexpected-contents task with 18-month-old infants, where the participants were familiarized with an experimenter reaching for blocks in three block boxes. Unbeknown to the experimenter, the fourth block box contained a spoon rather than a block. In this false-belief condition, when the experimenter expressed an interest in the contents of the fourth box, infants preferred to hand him a block rather than a spoon. By contrast, the reverse pattern of results was observed in the true-belief control, in which the experimenter knew what was inside the fourth box when he expressed an interest in its contents.

The 18-month-old infants in Buttelmann et al. (2014) were able to pass this modified unexpected-contents task. However, it must be noted that infants, on the one hand, and toddlers and children, on the other, have been set different standards for passing false-belief tasks. Whereas children are considered to pass a false-belief task only when they perform above chance (Wellman et al., 2001), infants are only expected to show a differential performance in false-belief and true-belief conditions. In the case of Buttelmann et al. (2014), there was a significant difference in infants' preference for the block between the false-belief condition (12 of 18) and the true-belief control (6 of 18), which was interpreted by the authors as evidence of infants' sensitivity to the experimenter's beliefs. However, these infants performed at chance in the false-belief condition (67% success rate), which would not be considered as passing the task in false-belief studies with toddlers and children. The different standards of success used at different ages have recently been discussed in replication studies that questioned whether performing differently in critical and control conditions can be reliably interpreted as evidence of false-belief understanding (see Kulke, Reiß, Krist, & Rakoczy, 2017; Kulke, Von Duhn, Schneider, & Rakoczy, 2018; Yott & Poulin-Dubois, 2016).

Besides the lower standards for false-belief tasks with infants, the results of Buttelmann et al. (2014) are also subject to richer and leaner interpretations. According to the authors, infants' preference for the block in the false-belief condition is evidence of early false-belief understanding. However, as in the case of the Duplo task, the design of this task allows for a different strategy based on memory representations; the familiarization trials (in which the experimenter got three blocks out of the first three block boxes) may have strengthened the association between the experimenter and the blocks (a sort of "block preference" representation), which may have allowed infants to select the block over the spoon without necessarily needing to infer that the experimenter had a false belief about the contents of the fourth box. In the true-belief condition, on the other hand, infants' preference for the spoon could be based on an association between the experimenter and the spoon inside the fourth box (because the experimenter had witnessed the contents of that box prior to the test phase).

Testing the role of memory and inference in modified false-belief tasks

The aim of the current study was to investigate the extent to which 3- and 4-year-old children may rely on associative memory representations to pass an unexpected-contents task modeled on the infant task by Buttelmann et al. (2014) and the design of the Duplo task (Rubio-Fernández & Geurts, 2013, 2016). Most current accounts of early Theory of Mind do not discuss the possible associative and inferential processes that may be required to pass verbal false-belief tasks (e.g., Helming et al., 2014; Westra & Carruthers, 2017). Thus, Baillargeon et al. (2010) argued that both infants and young children are able to attribute false beliefs to others, but standard false-belief tasks are harder than the modified versions designed for infants because producing the correct answer to a false-belief question requires greater Executive Function than anticipating what the agent will do in a nonverbal task. It follows from this account that narrative versions of unexpected-contents tasks that lower the executive demands of the original paradigm should be easier for toddlers. However, Baillargeon et al. did not make specific predictions regarding the possible roles of associative memory and inferential processes in false-belief reasoning.

The two-system account put forward by Apperly and Butterfill (2009; see also Low, Apperly, Butterfill, & Rakoczy, 2016; Rakoczy, 2017) makes more specific predictions regarding these processes. According to these authors, humans possess two separate systems for belief reasoning: a simpler, evolutionary and ontogenetically ancient system that operates fast and independently from central cognitive resources (S1) and a more sophisticated system that develops later and operates more slowly, making substantial demands on Executive Control processes (S2). Crucially, each system operates with a different conceptual apparatus — S1 with "belief-like" states and S2 with actual beliefs. This means that the early-developing system can track an agent's perceptions but treat them as perceptual registrations (rather than beliefs), whereas the later-developing system can attribute beliefs to agents.

According to this account, 3-year-olds would rely on associative memory representations to pass verbal false-belief tasks, whereas 4-year-olds would succeed in those tasks by deriving belief inferences about the protagonist. In the current study, the S1–S2 account would predict that 3- and 4-year-olds would only be able to pass those versions of the unexpected-contents task where they are able to form a memory representation associating the protagonist with the expected contents of the box. Without such a memory trace, unexpected-contents tasks would be too demanding for children under 4 years of age.

Motivation for the current study

There is a growing literature on failed replications of false-belief studies with infants, young children, and even adults (see, e.g., Kulke et al., 2017, 2018; Powell, Hobbs, Bardis, Carey, & Saxe, 2018; Yott & Poulin-Dubois, 2016). In the case of the Duplo task, whose design motivated the current study, the results have been mixed. Rubio-Fernández and Geurts (2016) replicated their original findings in a follow-up study, and Białecka-Pikul, Kosno, Białek, and Szpak (2019) also recently replicated these findings with a group of 210 children who went from below-chance level in a modified Duplo task at age 3;0 (years;months) to above-chance level at age 3;6 (with large effect sizes of 4:0). Kammermeier and Paulus (2018), on the other hand, were able to only partially replicate the original findings by Rubio-Fernández and Geurts (2013). Hence, only their 4-year-olds reached above-chance performance in the Duplo task, with their 3-year-olds performing at chance. However, both age groups performed significantly better in the Duplo task than in two standard false-belief tasks (where 3-year-olds performed below chance and 4-year-olds performed at chance). This pattern of results replicates and extends the significant difference observed by Rubio-Fernández and Geurts (2013) between 3-year-olds' performance in the Duplo task and the Smarties task (a pattern also replicated by Białecka-Pikul et al., 2019).

Kammermeier and Paulus (2018) failed to acknowledge their partial replication of the results of Rubio-Fernández and Geurts (2013) and favored a negative interpretation of their own findings (e.g., reanalyzing their data by counting children's uncooperative responses as Theory of Mind failures). Kammermeier and Paulus's approach suggests a negative confirmation bias, countering the rich interpretations of infant Theory of Mind studies proposed by some research groups (for discussion, see Heyes, 2014; Ruffman, 2014). Kammermeier and Paulus (2018) applied a double standard when interpreting their mixed results. They motivated their study by predicting that if 3-year-olds understand false beliefs, they should pass the Duplo task, and indeed they interpreted their chance performance as evidence that children under 4 years of age do not show an early understanding of belief. However, when discussing the better performance observed with both 3- and 4-year-olds in the Duplo task relative to two standard false-belief tasks, Kammermeier and Paulus dismissed the Duplo task as a genuine test of false-belief understanding. In other words, 3-year-olds' chance performance in the Duplo task was used to question their Theory of Mind abilities, whereas their better performance in the Duplo task relative to two standard false-belief tasks was dismissed as uninterpretable (for further discussion, see Paulus & Kammermeier, in press; Rubio-Fernández, in press-a, in press-b).

It must be noted that the results of the Duplo task were mixed even in the original studies by Rubio-Fernández and Geurts (2013, 2016); of the six versions of the Duplo task that these authors designed by introducing subtle manipulations to the original paradigm, 3-year-olds were able to pass only two (an important finding that Kammermeier and Paulus (2018) did not acknowledge either). The fact that 3-year-olds' ability to pass the Duplo task is so sensitive to subtle task manipulations suggests that this ability is rather fragile and requires careful investigation. The results of a recent false-belief study by Papafragou, Fairchild, Cohen, and Friedberg (2017) support this view; whereas previous studies had shown that young children were better able to show sensitivity to an actor's false belief in a word-learning setting than in a standard false-belief task, Papafragou et al. showed that when both kinds of tasks were equated for inhibitory control, the advantage of the communicative task disappeared.

While highlighting the importance of carefully analyzing task demands when interpreting Theory of Mind findings, the results of Papafragou et al. (2017) should not be overinterpreted as a license to conflate different paradigms and challenge previous studies. For example, if researchers wish to independently evaluate whether the Duplo task is easier for young children than a standard change-of-location task (for a similar advantage, see Lewis et al., 1994), it is important that they stay faithful to the original protocols of these tasks; otherwise, null results are inconclusive. Thus, in a recent replication attempt with the Duplo task, Wenzel, Dörrenberg, Proft, Liszkowski, and Rakoczy (2018) observed chance performance with 3-year-olds in both the Duplo task and a standard change-of-location task. However, the tasks used by these researchers differed from the original paradigms in two crucial ways. First, in their Duplo task, the experimenter asked the child "Did Freddy see that we moved the bananas?" whereas the original prompt was "Did the girl see what happened?" This is a key methodological difference because the studies by Rubio-Fernández and Geurts (2013, 2016) showed that mentioning the target object during the test phase draws children's attention to the actual location of the object and compromises their performance.

Second, the standard false-belief task used as a control included a feature of the Duplo task that is not present in standard change-of-location tasks; after the protagonist left the scene, the child was asked (a) whether the protagonist could still see the experimenter and the child, and once the object was transferred to the second container, (b) whether the protagonist had seen the transfer of the object. Rubio-Fernández and Geurts (2013) introduced these prompts in the original Duplo task to ensure that children would not lose track of the protagonist's perspective during the narrative, and their second experiment showed that these prompts were actually critical for young children's success. Therefore, the introduction of these prompts in a supposedly "standard" false-belief task may have improved children's performance, further blurring any possible differences between the Duplo and control tasks.

What the results of the original Duplo studies and subsequent replication studies unequivocally show is that 3-year-olds' ability to pass a change-of-location task is rather limited and highly dependent on very specific task manipulations (see also Lewis et al., 1994). Therefore, the current study was conceived as a follow-up to previous studies with the Duplo task in that it tried to unconfound those processes that are necessary for 3-year-olds to succeed in a modified false-belief task. Rubio-Fernández and Geurts (2013) never interpreted their findings as unequivocal evidence that 3-year-olds understand false beliefs, and they were open to the possibility that young children may rely on low-level mechanisms to pass the Duplo task (Rubio-Fernández & Geurts, 2016; cf. Kammermeier & Paulus, 2018). In line with that open-interpretation approach, the current study tested the hypothesis put forward by Rubio-Fernández and Geurts (2016) as to how 3-year-olds may rely on memory representations associating the protagonist with the correct outcome in order to pass modified false-belief tasks. That would explain, for example, why for 3-year-olds to pass the Duplo task, they need to focus on the protagonist throughout the narrative, whereas drawing their attention to the target object compromises their performance (cf. Białecka-Pikul et al., 2019).

Experiment 1

Experiment 1 investigated the Theory of Mind abilities of 4-year-old children in a standard Smarties task (Perner et al., 1989) and a modified version (the Secretive Smarties task) highlighting the secrecy of the contents of the Smarties tube. Previous studies have shown that introducing an element of secrecy and deception in a false-belief narrative may help young children to perform better in

Theory of Mind tasks (Chandler, Fritz, & Hala, 1989; Sullivan & Winner, 1993; see also Wellman et al., 2001). Unlike in the standard version of the task, in the Secretive Smarties task the experimenter highlights that the real contents of the Smarties tube are a secret between the experimenter and the child and that "nobody else knows what is in here."

Passing the Secretive Smarties task should in principle be easier than passing the standard version because other people's ignorance of the contents of the tube is highlighted in the protocol. However, passing this task still requires inferring that if people do not know what is inside the tube, they will rely on their world knowledge to guess what is inside (i.e., they will rely on the tube label or the appearance of the container). It is possible, therefore, that 4-year-old children, who are on the verge of starting to pass standard false-belief tasks, may need only a subtle manipulation highlighting the secrecy of the contents of the tube to pass a Smarties task. This result would suggest that 4-year-olds are capable of deriving false-belief inferences so long as other people's ignorance is highlighted for them. On the other hand, it is also possible that even 4-year-olds may find the Secretive Smarties task to be difficult because of its high inferential demands. That is, even if children this age may understand that a secret is shared only by those who know about it, they might not be able to generate the correct response to the test question (i.e., infer what the expected contents of the tube would be for someone who does not share in the secret). Experiment 1, therefore, was used to establish whether the inferential component of the unexpected-contents task is taxing for preschool children.

Method

Participants

A total of 38 children were recruited from a local preschool in northern Spain. The preschool is part of a public primary school and serves middle-class families. The children were tested at the end of their first year of preschool. The group consisted of all the children in the two available classrooms who received permission to participate in the study. There were 15 girls and 23 boys in the group, and their mean age was 4;0 (range = 3;6–4;5).

Materials and procedure

Children were randomly divided into two groups and tested on one of two versions of the Smarties task. In the standard version, the experimenter shows the child a tube of Lacasitos (the Spanish brand of Smarties) and asks the child what is inside the tube. When the child responds (according to the expected contents of the tube), the experimenter tells the child that she had already eaten all the Lacasitos but has put pencils inside; the experimenter opens the tube to show its real contents to the child. The experimenter then closes the tube again and asks the child the standard test question: "If we call in another kid who has not seen my tube, and we ask him what is inside, what is he going to say?"

The Secretive Smarties task is similar to the standard version, the difference being that when the experimenter closes the tube, she tells the child that the contents of the tube is a secret (she lowers her voice) and nobody else knows what is inside. The test question was also modified to emphasize the secrecy of the contents of the tube: "If we call in another kid but we don't tell him our secret, what is he going to say is inside the tube?"

Results and discussion

In addition to standard frequentist tests, Bayes factors (BFs) were used to explicitly test the hypotheses of interest and provide a measure of effect size. Large BF values correspond to strong evidence in favor of a hypothesis, small BF values correspond to strong evidence against a hypothesis, and BF values around 1 reflect inconclusive evidence. If children lack any understanding of this task, we would expect their performance on this test to be at chance (H0: $p = .50$). If children make the correct inference, we would expect their performance to be a mixture of their competence and some performance issues: (H1) $p = \alpha + (1 - \alpha)/2$, where α is a free parameter reflecting performance issues. At low values of α, children will respond at chance; at high values of α, children will respond near perfectly on the task. If children make the incorrect inference (H2), we expect their behavior to reflect performance issues: $(1 - \alpha)/2$.

Informed by past studies with 3-year-olds (Białecka-Pikul et al., 2019; Rubio-Fernández & Geurts, 2013), a beta (43, 29) prior was placed on the performance parameter given that approximately 20% of a sample in a Smarties task tends to respond opposite to the dominant response. The performance parameter was fit using Stan (Stan Development Team, 2016, 2018). To calculate BF, we marginalized over the performance parameter using bridge sampling (Gronau & Singmann, 2017). A uniform prior was placed over the hypotheses of interest. Table 1 provides BF and credible intervals for α for all experiments in the study.

As shown in Fig. 1, of the 18 children tested in the standard Smarties task, 6 responded correctly (according to the expected contents of the tube: "Lacasitos" or "chocolates") and 12 responded incorrectly (according to the actual contents of the tube: "pencils" or "colors"). Children performed at chance in this version of the task ($p = .238$, two-choice binomial test, two-tailed, $BF_{01} = 1420$). Of the 20 children tested in the Secretive Smarties task, 1 child did not cooperate with the experimenter and was eliminated from the analyses. Of the remaining 19 children, 7 responded correctly and 12 responded incorrectly. Children were also at chance in the modified version of the Smarties task ($p = .359$, two-choice binomial test, two-tailed, $BF_{01} = 935$), performing comparably in both tasks ($p = .823$, chi-square test).

The results of Experiment 1 suggest that even at around 4 years of age, children are not able to pass an unexpected-contents task that heavily depends on the derivation of belief inferences, not even when the unexpected contents are presented as a secret between the experimenter and the child and other people's ignorance of the contents of the tube is highlighted. Because this type of manipulation has proved to be successful in change-of-location tasks (Wellman et al., 2001), the question remains as to why 4-year-olds' performance in the Smarties task does not improve when an element

Table 1
Estimated performance parameter α and Bayes factors for each task.

Task	Mean α	Credible interval	Bayes factor	Evidence strength
Smarties (E1)	.52	.42–.63	$BF_{01} = 1420$	Strong
Secret Smarties (E1)	.52	.42–.63	$BF_{01} = 935$	Strong
Narrative Smarties (E2a)	.53	.43–.64	$BF_{01} = 67$	Substantial
Standard Smarties (E2b)	.45	.36–.55	$BF_{01} = 10372195$	Strong
			$BF_{20} = 1966$	
Narrative Associative Smarties (E2b)	.61	.50–.72	$BF_{10} = 1966$	Strong
True Belief Control (E2c)	.52	.42–.63	$BF_{01} = 462$	Strong

Note. BF_{XY} denotes the support for Hypothesis X over Hypothesis Y. E1/2a/2b/2c, Experiment 1/2a/2b/2c.

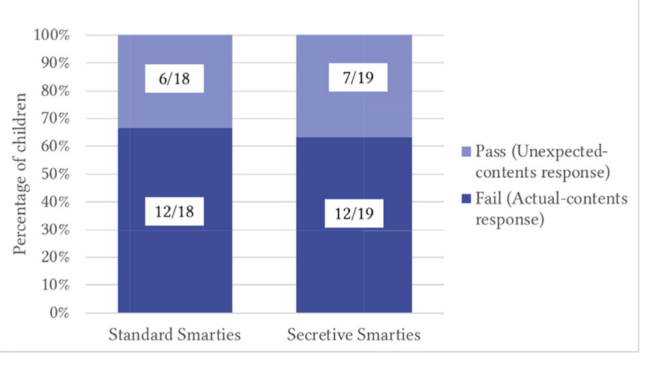

Fig. 1. Percentage of 4-year-old children who gave each type of response in the unexpected-contents tasks used in Experiment 1 (between-participants design).

of secrecy is introduced into the standard paradigm. A possible explanation, suggested by Rubio-Fernández and Geurts (2016), is that change-of-location tasks are easier than unexpected-contents tasks because children are able to form a memory trace associating the protagonist with the correct response (i.e., the original location of the object), whereas no such association is available in unexpected-contents tasks.

Experiment 2a

The hypothesis that associative memory representations may be important for young children's success in false-belief tasks was tested in Experiments 2a, 2b, and 2c. Here we report a longitudinal study with a group of 3-year-olds who were tested at 3-month intervals during their first year of preschool; a narrative/nonassociative Smarties task was used at Test Point 1, a standard Smarties task (used as a control) and a narrative/associative Smarties task were used at Test Point 2, and a true-belief control (modeled after the narrative/associative Smarties) was used at Test Point 3.

Experiment 2a introduced a new unexpected-contents paradigm that aimed to highlight the protagonist's perspective throughout the narrative (along the lines of the Duplo task) but still required participants to derive an inference about the protagonist's beliefs in order to select the correct response. This is the paradigm we dubbed the narrative/nonassociative Smarties task.

Method

Participants

A new group of 29 children was recruited from the same local preschool in northern Spain. The children were tested for the first time 3 months into their first year of preschool. The group consisted of all the children in the two available classrooms who received permission to participate. There were 15 girls and 14 boys, and their mean age was 3;6 (range = 3;0–3;11).

Materials and procedure

The narrative/non-associative Smarties task was modeled after the task designs by Buttelmann et al. (2014) and Rubio-Fernández and Geurts (2013). The materials are two traditional pencil boxes (showing pencils on the cover), one containing pencils and the other one containing drinking straws (same number and colors as the pencils), a wooden block (to prop the pencil boxes upright), a medium-size plastic figure of Winnie the Pooh, and a house (the size of the figure).

The two pencil boxes are placed vertically toward the left of a table, and the house is placed toward the right. The experimenter and the child are sitting next two each other, the experimenter facing the pencils and the child facing the house. The experimenter introduces the character Winnie de Pooh (which most Spanish preschoolers can recognize). The experimenter tells the child that Winnie loves drawing and asks whether the child likes drawing too. Then Winnie tells the child that he wants to show the child his new pencils (the experimenter puppeteers the figure using a high-pitched voice), takes the pencil box on the left, and empties it on the table. The experimenter comments how nice Winnie's new pencils are and puts them back in the box. Then she puts the pencil box back where it was. Winnie says that he is tired from drawing all day and wants to go and have a siesta in his little house. When the experimenter has put away the figure, she asks whether the child would like to look at the second box. The experimenter takes the pencil box on the right and empties it on the table. The experimenter acts surprised to discover that the box contains straws and tells the child, "But these are not pencils—these are straws!" The experimenter continues to act puzzled but then realizes, "I know what happened here; a naughty kid came in and took Winnie's pencils when Winnie was away!" Then, to draw the child's attention toward the protagonist (as in the Duplo task), the experimenter points to the house (where Winnie is supposed to be asleep) and adds in a low voice, "And poor Winnie didn't see the naughty kid take away his pencils! Oh-oh," with the experimenter looking worryingly at the child. The experimenter tidies away the straws and places the second box in its position. Then the experimenter points to the house and says, "Look, Winnie has woken up!" and brings the figure out again. Winnie tells the child that he wants to show the child his "collection," and the experimenter places the figure against the second pencil box, with its arms raised above its head. The expression "my collection" was chosen as neutral between the two possible contents of the box (the expected and the unexpected). The experimenter then tells the child, "You're a good kid. Why don't you help Winnie?" and produces a tray from under her chair holding a red pencil and a red straw (left–right position counterbalanced across children) and asks the child, "Why don't you help Winnie and give him what he is looking for?" If the child does not respond, the child is asked again, "What is Winnie looking for?"

It must be noted that, unlike in Buttelmann et al. (2014) where infants saw three blocks versus one spoon (potentially increasing the salience of the correct response in the critical condition), the children in the narrative/non-associative Smarties task saw the same number of pencils (first box) and straws (second box), making sure that the two responses to the test question were balanced.

Results and discussion

As shown in Fig. 2, of the 29 children in the study, 1 child failed to cooperate and was eliminated from the analyses. Of the remaining 28 children, 15 selected the pencil (expected contents, correct response) and 13 selected the straw (actual contents, incorrect response). Children in the narrative/non-associative Smarties task, therefore, were at chance (p = .851, two-choice binomial test, two-tailed, BF_{01} = 67).

The results of Experiment 2a suggest that 3-year-old children are not able to pass an unexpected-contents task that relies on the derivation of belief inferences without the support of associative memory representations. These findings are in contrast to 3-year-olds' successful performance in the Duplo task. The results of Rubio-Fernández and Geurts (2013, 2016) suggest that 3-year-olds are able to pass a change-of-location false-belief task provided that (a) they are able to focus on the protagonist throughout the narrative and (b) the incorrect response is not made more salient than the correct response (by drawing children's attention to the target object during the test phase). The design of the narrative/non-associative Smarties task used in Experiment 2a was comparable to that of the Duplo task in both of these features. However, the 3-year-olds in Experiment 2a performed at chance, whereas children of the same age in the Duplo task were above chance.

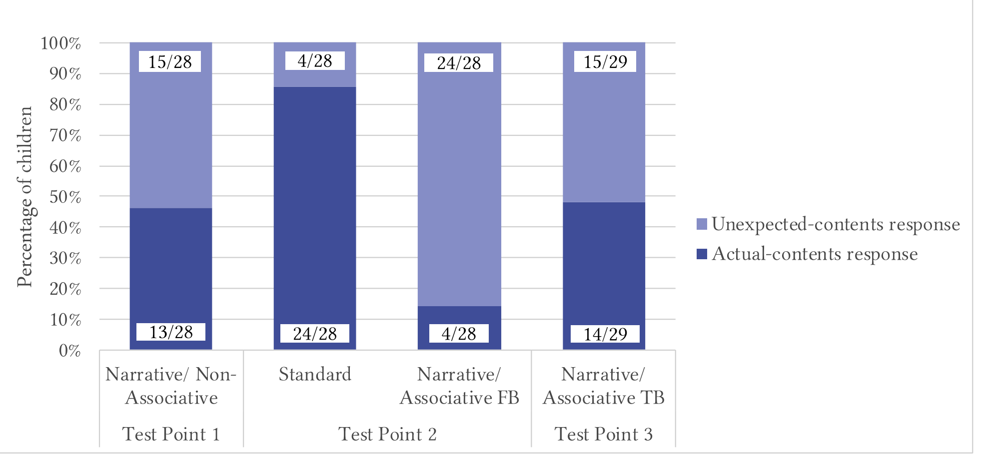

Fig. 2. Percentage of 3-year-old children who gave each response type in the four versions of the unexpected-contents tasks used in Experiment 2a (Narrative/Non-Associative Smarties—Test Point 1), Experiment 2b (Standard Smarties and Narrative/Associative Smarties—Test Point 2), and Experiment 2c (Narrative/Associative True-Belief control—Test Point 3). Three months lapsed between each test point (within-participants design). Note that, unlike in the three false-belief tasks, the correct response in the true-belief control was the actual contents of the box. FB, false belief; TB, true belief.

Experiment 2b

A possible key difference between Experiment 2a and the Duplo task is that in the latter children were able to form a memory representation of the protagonist placing the object in the original location (a representation that forms the basis for the protagonist's false belief). By contrast, children in the narrative/non-associative Smarties task needed to infer that the protagonist was looking for pencils when he tried to open the second box containing straws. To investigate whether this difference is critical for early success in verbal false-belief tasks, Experiment 2b employed an associative version of the unexpected-contents task used in Experiment 2a, where the same group of 3-year-old children was able to form a memory representation associating the protagonist with the expected contents of the box. To control for the possibility that children may have developed the ability to pass verbal false-belief tasks during the 3 months between Experiment 2a and Experiment 2b, the latter included a standard Smarties task that served as a baseline.

Method

Participants

The same group of children (minus one who was sick, $N = 28$) was recruited from the same local preschool in northern Spain. The children were tested for the second time 6 months into their first year of preschool. The group consisted of 14 girls and 14 boys, and their mean age was 3;9 (range = 3;3–4;2).

Materials and procedure

To control for the possibility that children in the study may have developed the ability to pass false-belief tasks during the previous 3 months, they were first tested on a standard Smarties task using the same protocol described in Experiment 1. The only difference was that the children in this group were not so familiar with Lacasitos (the Spanish Smarties), and so the set of materials used for the standard unexpected-contents task was a box of dinosaur biscuits (showing dinosaur biscuits on the cover) that contained crayons. All children were first tested on the standard Smarties task.

On the second day of testing, children were presented with the narrative/associative Smarties task. To avoid the fact that children may remember any details of the non-associative version, a new set of materials was used in this version of the task: two boxes of colorful plasters (showing plasters on the cover), one empty and one containing plasters, a set of ice-cream spoons (same number, size, and colors as the plasters in the second box), a wooden block (to prop the boxes upright), a Duplo figure, and a Duplo house. The materials were placed in the same positions as in Experiment 2a.

The experimenter starts by introducing the Duplo girl and explains that she loves plasters and asks the child whether the child likes plasters too. Then she explains that the Duplo girl likes plasters so much that she has used an entire box, and she shows that the box to the left is empty. The Duplo girl then says (in a high-pitched voice) that, luckily, she still has some plasters in the other box, and she empties its contents on the table. This is the point in the narrative where children are able to form a memory representation associating the protagonist with the expected contents of the box. The experimenter praises the colorful plasters and then tidies them away and puts the second box back where it was. Then the Duplo girl says she wants to go and have a siesta and goes into her house. The experimenter, pointing to the house in a secretive manner, asks the child, "Can the Duplo girl see me from over there? No, she can't!" and with an expression suggesting connivance, she empties the second plaster box on the table, produces a bunch of ice cream spoons from her pocket, and says, "Look, I have some colorful spoons here!" The experimenter then swaps the two contents with an amused mischievous attitude, asking the child to check that the Duplo girl has not seen what happened (as in the Duplo task). The Duplo girl then comes out from her house and tells the child that she wants to show him her collection, and the experimenter places the Duplo figure against the second box with raised arms. The experimenter then tells the child, "You're a good kid. Why don't you help the girl?" and produces a tray from under her chair holding a green plaster and a green spoon (left–right position counterbalanced across children) and asks the child, "Why don't you help the girl and give her what she is looking for?" As in Experiment 2a, the two possible responses to the test question are equally salient, with children having seen one set of plasters and one set of spoons.

Results and discussion

As shown in Fig. 2, of the 28 children in the group, 4 passed the standard Smarties task (responding according to the expected contents of the box: "biscuits" or "dinosaurs"), whereas 24 failed (responding according to the actual contents of the box: "colors"' or "crayons"). Children in this test performed significantly below chance ($p < .001$, two-choice binomial test, two-tailed, $BF_{20} = 1966$).

In the narrative/associative Smarties task, 24 children passed the task (giving the Duplo girl a plaster: the expected contents of the box) and 4 children failed (giving the Duplo girl a spoon: the actual contents of the box). Children in this version of the Smarties task performed significantly above chance ($p < .001$, two-choice binomial test, two-tailed, $BF_{10} = 1966$). A McNemar test with continuity correction revealed a significant difference between the standard Smarties and the narrative/associative Smarties task, $\chi^2(1, N = 27) = 18.05$, $p < .001$.

The results of Experiments 2a and 2b were also compared. McNemar tests with continuity correction revealed a significant difference between the standard Smarties and the narrative/non-associative Smarties task, $\chi^2(1, N = 26) = 7.692$, $p = .006$, and a significant difference between the associative and non-associative versions of the narrative Smarties task, $\chi^2(1, N = 26) = 4.267$, $p = .039$.

The results of Experiment 2b suggest that the 3-year-old children tested in this longitudinal study had not undergone enough Theory of Mind development during the 3 months that separated Experiment 2a and Experiment 2b as to be able to pass a standard false-belief task, performing significantly below chance in the Smarties task that was used as a baseline. In fact, comparing the standard Smarties with the narrative/non-associative version that was used 3 months earlier, these children performed significantly better at the earlier testing point, when they were at chance. This difference suggests that the narrative elements of the modified Smarties task (i.e., the constant focus on the protagonist and the balanced responses to the test question) helped children to perform better than in the standard false-belief task. This is in line with what Rubio-Fernández and Geurts (2013, 2016) observed with the Duplo and Smarties tasks and what Lewis et al. (1994) predicted in their earlier work on narrativity and false-belief tasks.

The children in Experiment 2b performed significantly above chance in the narrative/associative Smarties task, also performing significantly better than in the standard Smarties task and the non-associative version of Experiment 2a. These results suggest that allowing 3-year-olds to form a memory representation associating the protagonist with the expected contents of the box is key for their successful performance in unexpected-contents tasks.

Experiment 2c

Even if 3-year-old children were able to pass the narrative/associative Smarties task, the question remained as to whether they did this by bearing in mind the protagonist's preference for the original contents of the box (as explained at the start of the task, when children were told that the protagonist loved plasters) or by attributing to the protagonist a false belief about the contents of the plaster box. This argument is related to the lean interpretation of Buttelmann et al.'s (2014) findings, whereby infants were able to develop a stronger preference for the expected contents of the last box by seeing the actor retrieve the same content from the previous three boxes. To tease apart these two possible strategies (agent's preference vs. false-belief attribution), the last experiment of this study employed a true-belief version of the narrative/associative Smarties task, where the protagonist agreed to exchange the contents of the box with the experimenter prior to the test phase.

If children rely on the protagonist's preference for the original contents of the box, they would select that object when trying to help the protagonist. By contrast, if children assume that the protagonist is looking for the current contents of the box, they would help her to get that object. The latter pattern of results would support a rich interpretation of the results of Experiment 2b, whereby children understand that the protagonist is looking for what she thinks is inside the box (which was mistaken in Experiment 2b but is correct in Experiment 2c). That means that 3-year-old children in the true-belief control should show a reliable preference for the current contents of the box, performing significantly different from the false-belief version used in Experiment 2b.

Note that here we are distinguishing two factors that could affect children's performance in this task. The first is whether the child could form a memory trace associating the protagonist with the original contents of the box. This feature distinguished the narrative/non-associative and narrative/associative Smarties tasks. The second factor is the protagonist's verbal expression of a preference for the expected contents of the box (e.g., Winnie's liking of pencils or the Duplo girl's liking of plasters). This is a common element to the two narrative Smarties tasks used in Experiments 2a and 2b and was tested in the true-belief control in Experiment 2c.

Method

Participants

The same group of 29 children was recruited from the same local preschool in northern Spain. The children were tested for the third time 9 months into their first year of preschool. The group consisted of 15 girls and 14 boys, and their mean age was 4;0 (range = 3;6–4;5).

Materials and procedure

Because passing a true-belief task does not require false-belief understanding, we did not use a standard false-belief control in Experiment 2c. To avoid the fact that the children may remember their responses to the task in Experiment 2b, Experiment 2c used the same set of materials as in Experiment 2a but with a similar script as in Experiment 2b (because this version is supposed to be a true-belief control for the narrative/associative task). The child is told that Winnie the Pooh loves pencils but has used all the pencils in the first box (the experimenter demonstrates that the box to the left is empty). Winnie adds that, luckily, he has some pencils in the second box. When Winnie empties the contents of the second pencil box on the table, the experimenter asks him whether he would like to trade with her, and she produces a set of straws from her pocket. Winnie agrees to trade his pencils for the straws, and the experimenter swaps the two contents and puts the second box in its place. Winnie then says that he wants to have a siesta and goes into his house to sleep. To make the task equally long as previous versions, the experimenter asks whether the child likes Winnie's house. When Winnie wakes up, he says he wants to show the child his collection, and the experimenter places the figure against the second box with raised arms. As in previous versions of the task, the experimenter then tells the child, "You're a good kid. Why don't you help Winnie?" and produces a tray from under her chair holding a red pencil and a red straw (left–right position counterbalanced across children) and asks the child, "Why don't you help Winnie and give him what he is looking for?"

Results and discussion

As shown in Fig. 2, of the 29 children in Experiment 2c, 14 selected the straw (responding according to the actual contents of the box: true-belief response) and 15 selected the pencil (responding according to the protagonist's original preference for that object). The children in this true-belief control, therefore, were at chance between the two responses ($BF_{01} = 462$). A McNemar test with continuity correction revealed a significant difference between the narrative/associative Smarties task in Experiment 2b and the true-belief version in Experiment 2c, $\chi^2(1, N = 26) = 6.750$, $p = .009$.

Children's preference for the actual contents of the box was significantly different in the false-belief and true-belief versions of the narrative/associative Smarties task used in Experiments 2b and 2c. In the infant Theory of Mind literature, such a difference would be understood as reliable sensitivity to the protagonist's beliefs (see Buttelmann et al., 2014). However, when trying to interpret children's responses in the false-belief condition, the results of the true-belief control reveal a mixed picture, with about half of the children responding according to the protagonist's preference for the previous contents of the box (as explained verbally at the beginning of the story) and the other half understanding that the protagonist wanted to get the actual contents of the box. The control condition, therefore, suggests that 3-year-olds' successful performance in the narrative/associative Smarties task used in

Experiment 2b cannot be unequivocally interpreted as evidence of their false-belief understanding because at least some of the children who passed the task may have done so by relying on the protagonist's preference for the original contents of the box (rather than by attributing to her a false belief about the current contents of the box).

General discussion

This study investigated the extent to which 3- and 4-year-old children may rely on associative memory processes to be able to pass verbal false-belief tasks. In the Duplo task, for example, young children are able to predict the correct outcome by continuously focusing on the protagonist through a change-of-location narrative (Rubio-Fernández & Geurts, 2013). However, focusing exclusively on the protagonist (and not on the target object) may allow 3-year-olds in the Duplo task to rely on a memory representation associating the protagonist with the original location of the object rather than necessarily deriving an inference about the protagonist's false belief (Rubio-Fernández & Geurts, 2016).

Passing standard unexpected-contents tasks requires deriving a belief inference given that the correct response (corresponding with the expected contents of the box) is not associated with the agent. The results of Experiment 1 suggest that deriving belief inferences is hard even for 4-year-old children, who perform comparably in a standard Smarties task and a modified version highlighting other people's ignorance of the actual contents of the tube (what we called the Secretive Smarties task). In both versions of this unexpected-contents task, 4-year-olds performed at chance even though introducing an element of secrecy and deception has previously been observed to help young children perform better in change-of-location false-belief tasks (Wellman et al., 2001). The results of Experiment 1 indirectly support Rubio-Fernández and Geurts's (2016) argument that unexpected-contents tasks may be harder than change-of-location tasks, at least for young children.

The results of the longitudinal study reported in Experiments 2a, 2b, and 2c suggest that 3-year-old children perform significantly better in a narrative version of the Smarties task (where they are allowed to continuously focus on the protagonist throughout the narrative and the two possible responses to the test question are equally salient) than in a standard Smarties task. These results support the view that standard false-belief tasks pose extra executive demands relative to the false-belief tasks adapted for infants (Baillargeon et al., 2010) and that features of task design that preserve the continuity of perspective tracking are important for 3-year-olds to show improved performance in verbal false-belief tasks (Rubio-Fernández & Geurts, 2013, 2016; for similar proposals, see Lewis, 1994; Lewis et al., 1994).

The 3-year-olds in this longitudinal study reached above-chance performance in a narrative version of the Smarties task that allowed them to form a memory representation associating the protagonist with the expected contents of the box, performing significantly better than in the non-associative version (where they were at chance) and the standard Smarties control (where they were significantly below chance). This pattern of results supports the two-system account put forward by Apperly and Butterfill (2009), according to which infants and young children are able to pass false-belief tasks by relying on memory representations (or *registrations*, in their terminology). In the case of the modified Smarties task used in Experiment 2b, 3-year-olds would have been able to associate the protagonist with the original contents of the box because the protagonist showed the contents to the child before she left the scene. This memory association seems to be critical for 3-year-olds to pass an unexpected-contents task, as the two-system account would predict.

While supporting the predictions of the S1–S2 account, it is important to acknowledge that the results of this study do not provide evidence that 3-year-olds are not able to attribute false beliefs to others. It is in principle possible that 3-year-olds may need to first form a mental representation of what the protagonist believes is inside the container before they can reason counterfactually about what she may be looking for when reaching into the container later in the narrative. Such an interpretation of our findings would be in line with Lewis et al.'s (1994) account of 3-year-olds' emerging Theory of Mind.

However, the results of the true-belief control suggest that when children selected the expected contents of the box in the false-belief tasks, they may have been relying on the protagonist's preference for that object given that some of the children did that in the true-belief task (where the protagonist knew what was inside the box). In other words, rather than selecting the response that corresponded with the actual contents of the box, about half of the children in the true-belief control responded according to the original contents of the box, which was introduced in the story as the protagonist's favorite.

Three conclusions can be unambiguously drawn from the results of this study. First, it is important to examine the specific task demands of modified and standard false-belief tasks before concluding that passing those tasks reveals false-belief attribution (or that failing those tasks reveals an inability to attribute false beliefs to others). Second, narrative versions of unexpected-contents tasks where children can continuously track a protagonist's perspective are easier than standard versions where children need to derive belief inferences to arrive at the correct response. In the standard Smarties task, for example, children must understand not only that other children may be ignorant of the actual contents of the tube but also that they will respond to the test question according to their world knowledge (i.e., that Smarties tubes normally contain chocolates). This inferential reasoning may be too difficult for 3-year-olds regardless of their Theory of Mind development, as suggested by the better performance observed in the two narrative versions of the Smarties task used in this study (see also Freeman & Lacohée, 1995; Mitchell & Lacohée, 1991).

Finally, preserving the continuity of the process of perspective tracking does not seem to be enough for 3-year-olds to pass an unexpected-contents task. At that age, children need to be able to form a memory representation associating the protagonist to the original contents of the box in order to select the expected contents during the test phase. Even in a true-belief control, where the protagonist and the child shared the same perspective, children often responded according to the protagonist's preference for the original contents of the box (rather than the protagonist's knowledge of the contents of the box). This pattern of responses suggests that when 3-year-olds passed the modified unexpected-contents task, they may have been sensitive to the protagonist's preferences and desires and not only (or not necessarily) to the protagonist's beliefs.

Acknowledgments

This research was supported by a Young Research Talent Grant from the Research Council of Norway (230718), and the author gratefully acknowledges this funding. Special thanks go to Francis Mollica for his help with the statistical analyses.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2018.08.011.

References

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

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 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.

Białecka-Pikul, M., Kosno, M., Białek, A., & Szpak, M. (2019). Let's do it together! The role of interaction in false belief understanding. *Journal of Experimental Child Psychology, 177*, 141–151.

Buttelmann, D., Over, H., Carpenter, M., & Tomasello, M. (2014). Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of Experimental Child Psychology, 119*, 120–126.

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language, 28*, 141–172.

Chandler, M. J., Fritz, A. S., & Hala, S. M. (1989). Small-scale deceit: Deception as a marker of two-, three- and four-years-olds' early theories of mind. *Child Development, 60*, 1263–1277.

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

Gronau, Q. F., & Singmann, H. (2017). Bridgesampling: Bridge sampling for marginal likelihoods and Bayes factors. R package Version 0.4-0. Retrieved from <https://CRAN.R-project.org/package=bridgesampling>.

Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences, 18*, 167–170.
Heyes, C. (2014). False belief in infancy: a fresh look. *Developmental Science, 17*(5), 647–659.
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. (2017). How robust are anticipatory looking measures of theory of mind? Replication attempts across the life span. *Cognitive Development, 46*, 97–111.
Kulke, L., Von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychological Science, 29*, 888–900.
Lewis, C. (1994). Episodes, events, and narratives in the child's understanding of mind. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 457–480). Hillsdale, NJ: Lawrence Erlbaum.
Lewis, C., Freeman, N. H., Hagestadt, C., & Douglas, H. (1994). Narrative access and production in preschoolers' false belief reasoning. *Cognitive Development, 9*, 397–424.
Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives, 10*, 184–189.
Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology, 30*, 1–13.
Mitchell, P., & Lacohée, H. (1991). Children's early understanding of false belief. *Cognition, 39*, 107–127.
Papafragou, A., Fairchild, S., Cohen, M. L., & Friedberg, C. (2017). Learning words from speakers with false beliefs. *Journal of Child Language, 44*, 905–923.
Paulus, M., & Kammermeier, M. (in press). How to deal with a failed replication of the Duplo task? A response to Rubio-Fernandez. *Cognitive Development*.
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.
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). In defense of a developmental dogma: Children acquire propositional attitude folk psychology around age 4. *Synthese, 194*, 689–707.
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.
Rubio-Fernández, P. (in press-a). What do failed (and successful) replications with the Duplo task show? *Cognitive Development*.
Rubio-Fernández, P. (in press-b). Trying to discredit the Duplo task with a partial replication: Reply to Paulus & Kammermeier (2018). *Cognitive Development*.
Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review, 34*, 265–293.
Stan Development Team (2016). StanHeaders: Headers for the R interface to Stan. Retrieved from http://mc-stan.org.
Stan Development Team (2018). RStan: The R interface to Stan. Retrieved from http://mc-stan.org.
Sullivan, K., & Winner, E. (1993). Three-year-olds' understanding of mental states: The influence of trickery. *Journal of Experimental Child Psychology, 56*, 135–148.
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.
Wenzel, L., Dörrenberg, S., Proft, M., Liszkowski, U., & Rakoczy, H. (2018, May). Children's understanding of aspectuality in an action-based false-belief task. Poster presented at the *X Dubrovnik conference on cognitive science: Communication, pragmatics and theory of mind*, Dubrovnik, Croatia.
Westra, E., & Carruthers, P. (2017). Pragmatic development explains the theory-of-mind Scale. *Cognition, 158*, 165–176.
Yott, J., & Poulin-Dubois, D. (2016). Are infants' theory of mind abilities well integrated? Implicit understanding of intentions, desires, and beliefs. *Journal of Cognition and Development, 17*, 683–698.