Infants' false belief understanding: A non-replication of the helping task

Cristina Crivello*, Diane Poulin-Dubois

Concordia University, 7141 Sherbrooke Street West, Montréal, Québec, H4B 1R6, Canada

ARTICLE INFO

ABSTRACT

False belief understanding in infancy can be measured through helping tasks. However, there has been controversy surrounding the interpretation of such data. In order to better understand the nature of infants' false belief understanding, replication of the original findings is critical. The objective of the present study was to conduct a conceptual replication of the helping task designed by Buttelmann, Carpenter, and Tomasello (2009). In Experiment 1, 41 18-month-olds were tested using the false belief task. A larger sample (n = 97) was tested in Experiment 2 with this task in order to increase statistical power. Additionally, 33 infants were tested on a true belief task. Results from both experiments failed to replicate those from the original study. The discussion addresses the potential reasons for this lack of replication, as well as the implication of these findings regarding the debate on the interpretation of infants' behavior in this prompted-action task.

1. Introduction

Theory of mind is defined as the ability to explain and predict others' behavior by attributing beliefs, knowledge, intentions, and desires (Wellman, 2014). Much of the research on children's theory of mind has been on false belief understanding, which refers to the process of recognizing that others may have beliefs that are different from reality (Schneider, Slaughter, & Dux, 2015). Although early research suggested that children pass a false belief task only by the age of five (Wellman, Cross, & Watson, 2001), research conducted over the last decade has provided evidence for false belief understanding in infancy (see reviews by Baillargeon, Scott, & He, 2010; Sodian, 2011).

Understanding of false belief in infancy has mainly been demonstrated with violation of expectation and anticipatory looking paradigms (also known as spontaneous-response tasks), which are tasks based on infants' spontaneous looking responses, as opposed to prompted verbal responses (Baillargeon et al., 2010; Yott & Poulin-Dubois, 2016). These paradigms measure implicit false belief understanding, in which "implicit" refers to a spontaneous, non-elicited behavior that does not require selecting a response or holding information in working memory (Baillargeon, Scott, & Bian, 2016). In an implicit false belief task, demands are minimal, which allow infants to demonstrate their competence with minimal language and executive function skills (Baillargeon et al., 2010). In contrast, explicit (also known as elicited-response) false belief tasks refer to tasks that require verbal answers regarding the informant's false belief (Baillargeon et al., 2010). Previous research has found that both children and adults have implicit and explicit false belief understanding, whereas infants only have implicit false belief understanding (Apperly & Butterfill, 2009). According to Apperly and Butterfill (2009), the development of implicit and explicit theory of mind is independent from one another. Specifically, they argue

that there are two systems to belief reasoning: one efficient system which is automatic, fast, and does not require much cognitive abilities (i.e., implicit) and one flexible system which is slow and does require language and executive function abilities (i.e., explicit). Implicit false belief is presumed to develop in infancy (e.g., Kovács, Téglás, & Endress, 2010; Onishi & Baillargeon, 2005) and remains stable across the lifespan (Apperly & Butterfill, 2009). In contrast, explicit false belief is presumed to develop later than implicit false belief, at approximately four years of age (Baillargeon et al., 2010).

Although implicit false belief tasks are frequently used in infancy, the rich interpretation of infants' behaviors in these tasks is currently a topic of debate (e.g., Heyes, 2014; Ruffman, 2014), and low pass rates as well as recent failures to replicate the original findings suggest a lack of robustness (Burnside, Ruel, Azar, & Poulin-Dubois, 2017; Dörrenberg, Rakoczy, & Liszkowski, 2017; Kulke & Rakoczy, 2017; Poulin-Dubois, Polonia, & Yott, 2013; Poulin-Dubois & Yott, 2017; Sodian et al., 2016; Thoermer, Sodian, Vuori, Perst, & Kristen, 2012; Yott & Poulin-Dubois, 2016). The criticisms against the VOE paradigm have led researchers to develop interactive tasks, also known as helping paradigms (Buttelmann, Carpenter, & Tomasello, 2009; Buttelmann, Over, Carpenter, & Tomasello, 2014; Knudsen & Liszkowski, 2012a; Knudsen & Liszkowski, 2012b; Southgate, Chevallier, & Csibra, 2010). The classic procedure involves having the child help the experimenter by informing (i.e., pointing) or retrieving an object for him/her, which requires the child to understand that the experimenter has a false belief about the location of the object (Poulin-Dubois & Yott, 2017). According to Baillargeon and colleagues (2015), the helping task is a hybrid between spontaneous- and elicited-response tasks, as a response selection is required, but no response inhibition is needed. Thus, there are fewer processing demands than explicit, elicited-response tasks, but more processing demands than implicit, spontaneous-response tasks, such as VOE paradigms.

Buttelmann and colleagues (2009) were the first to develop an active behavioral measure based on helping to assess infants' understanding of false and true belief. In this study, children were randomly assigned to a true belief or false belief condition. In both conditions, children observed an experimenter switch the location of a toy from one box to the other. Another experimenter (the protagonist) observed this switch in the true belief condition, whereas he did not in the false belief condition. Following the switch, the protagonist attempted but failed to open the box in which the toy was originally placed. Children in the true belief condition were expected to help the experimenter open the empty box, as they should have inferred that the experimenter did not try to open the empty box to retrieve the toy as he observed the toy being moved. In contrast, children in the false belief condition were expected to open the box where the toy was presently located in order to retrieve it for the experimenter. This behavior would be aligned with children's understanding that the experimenter had a false belief about the location of the toy and inferred that the experimenter wanted to retrieve it from the location he thought the toy remained. The authors demonstrated that 75% of 2.5-year-olds correctly opened the empty box the experimenter had tried to open in the true belief condition, whereas 83.3% of the children correctly opened the box containing the toy in the false belief condition. These results were successfully replicated in a second experiment with 18-month-olds, demonstrating a pass rate of 84% in the true belief condition and 72% in the false belief condition. Although similar results were obtained for the 16-month-old group, infants' performance in the true belief condition was not above chance. Of importance is the exceptionally high attrition rate in this study (54%). Specifically, 119 of the 16- and 18-month-olds were excluded from the analyses, resulting in a final sample of 50 infants in each group.

In addition to the high attrition rate, other criticisms have been raised regarding Buttelmann and colleagues' (2009) false belief task. One argument is that children may not be using the experimenter's false belief to decide which box to open, but may simply be influenced by the broader social situation (i.e., playing a trick on the experimenter) (Allen, 2015). In other words, playing a trick on the experimenter by hiding the toy makes the toy more salient and therefore increases the children's expectancy that the experimenter will search for his toy (Allen, 2015). This trickery only occurs in the false belief condition, making it an important difference across conditions (Allen, 2015). In Buttelmann and colleagues' (2009) study, it was noted that before E2 tried to open the box that originally had his toy inside, approximately seven participants attempted to inform the experimenter (e.g., by pointing) that the toy was no longer in the original box, whereas only one child attempted to inform the experimenter of this in the true belief condition (Allen, 2015). Children should not have been able to infer the experimenter's goal since he had not yet tried to open the box where he had initially placed his toy. Therefore, Allen (2015) concluded that there must be an additional variable, such as the social situation of playing a trick on the experimenter, which led children to assume that the experimenter wanted his toy. In fact, Allen (2015) investigated preschoolers' helping behavior with the Buttelmann and colleagues' false belief task, and included two control conditions that involved no opportunity to play a trick on the experimenter. The results demonstrated that the preschoolers did not use the experimenter's false belief in order to help, but instead used the broader social situation. In agreement, Perner (2014) argued that the conspiratorial context of the false belief condition relates to hide and seek, which could explain why infants try to help the experimenter find his toy. In a hide and seek game, children expect individuals to initially look in the wrong location before finding the person, and they enjoy providing hints as to where the hidden person is located (Perner, 2014). Taken together, researchers have argued that children's responses on an interactive false belief task could be driven by social context (i.e., playing a trick, hide and seek), and not necessarily by children's ability to understand others' mental states.

Furthermore, Priewasser, Rafetseder, Gargitter, and Perner (2017) have suggested an alternative explanation from a teleological perspective. First, a strict replication of Buttelmann and colleagues' false and true belief tasks was conducted in Study 1. Although the researchers could not replicate the true belief task (helping E2 to open the empty box), they replicated the original pattern of responses for the false belief task (helping E2 open the box with the toy) as well as a significant difference in the box opened across the false and true belief conditions. In Study 2, an alternative interpretation to Buttelmann and colleagues' findings was tested by introducing a third box in the false and true belief task to test the teleological interpretation. The new false belief condition was identical to the original condition except that E2 tried to open this third, empty box, instead of trying to open the box where the toy was before the transfer. In the false belief condition, the authors hypothesized that infants would recognize that E2 was interested and emotionally engaged with the toy. Therefore, according to a teleological view, when E2 tries to open the third box, infants should

display a tendency to help and open the box where the toy is currently located, without the need to infer E2's false belief. In contrast, the mentalistic view predicts that infants should open the third empty box, as they should infer that this is the box that E2 wants to open since she knows that the toy is not located there. The results showed that infants directed the adult to the box with the toy in this new false belief task, providing evidence for the teleological perspective and not the mentalistic one. Therefore, the findings of this study strongly challenge Buttelmann and colleagues' argument that young children's performance on this task is based on attributing a false belief to the experimenter. Instead, it is argued that young children infer what the experimenter wants (i.e., the toy) and helps her achieve her goal of getting the toy, without reasoning about the experimenter's mental state.

Given the current debate about the interpretation of helping behaviors in the prompted-action, helping paradigm, both strict and conceptual replications of this task, as well as more control conditions, are needed to assess the robustness of infants' false belief understanding. Thus, the objective of the study was to examine the performance of two samples of 18-month-olds in a conceptual replication of the original study by Buttelmann and colleagues (2009), as slight methodological changes were made to the original procedure. These methodological changes were made in order to minimize attrition, as a large number of children did not demonstrate helping behavior in the original study. In Experiment 1, the false belief task was administered to 41 18-month-old infants seated at a table, instead of being administered on the floor as in the original study. In Experiment 2, a much larger sample of 18-month-old infants was tested ($n = 97$) on the false belief task with a larger distance between the boxes and the infant. As in the original study, an additional sample ($n = 33$) was tested in a true belief condition.

2. Experiment 1

2.1. Method

2.1.1. Participants

The final sample consisted of 41 infants ($M_{age} = 18.40$ months, $SD = 0.69$; range = 17.2–20.3; 19 males, 22 females). Nine participants were excluded due to fussiness ($n = 6$), parental interference ($n = 1$) and experimenter error ($n = 2$). Seven additional participants were excluded due to reasons specific to the task (described below). Infants were recruited through birth lists that were made available from a governmental health agency. In order to participate, participants were required to be exposed to English or French, and to be free of any auditory or visual impairment.

2.1.2. Materials

In the false belief task (Buttelmann et al., 2009), two wooden boxes, measuring $30 \times 30 \times 30$ cm, were used. One box was painted green and the other box was painted orange. Both boxes had lids with a hole (2 cm diameter), which could be locked with wooden pins (5 cm long) by insertion into the opening. A toy caterpillar was also used in this task.

2.1.3. Procedure

Infants were first familiarized with the environment and the experimenters. During this time, the caregiver signed the consent form. Parents were given $20 as financial compensation, and infants were given a small gift as well as a certificate of merit.

2.1.3.1. False-belief task.
Infants' ability to understand that others may have different beliefs was assessed using an interactive task adapted from Buttelmann and colleagues (2009). In order to maximize children's cooperation, a slight change in the methodology was introduced: the task was completed at a table instead of on the floor, ensuring that the children did not have to move toward the display. This modification was also introduced to minimize changes in the set-up, as other tasks were administered to infants at a table as part of a larger study. It is important to note that no order effect was observed. The two wooden boxes were positioned on the table, approximately 15 cm in front of the infant. At the beginning of the task, the second experimenter (E2) said to the infant, "Wow, look at these boxes!", as she opened each box and demonstrated to the infant that both boxes were empty. E2 then announced to the infant that she was going outside to get a toy. During the time that E2 was away, the first experimenter (E1) taught the infant how to lock and unlock the set of boxes with wooden pins. The infant was asked to unlock and lock each box twice. Following this, E2 returned with a toy caterpillar and played with the infant for 90 s or until the infant became uninterested in the toy. E2 then told the infant that she would put her toy in one of the boxes, and placed the toy inside as the infant observed. Once the toy was placed inside one of the boxes, E2 announced, "Oops! I forgot my keys outside. I'll be right back!", and left the room. E1 then said to the infant, "Let's play a trick on (E2's name). I'll take the toy from this box and put it in this box here. Shhhh!", as she removed the toy from its current location and placed it in the other box. She then said, "Now let's lock the boxes.", as she inserted the pins into both boxes. Next, E2 returned and stood in front of the boxes and said, "So.", then attempted to open the box in which she had initially placed her toy. Once E2 failed to open the box, she turned toward the infant and stated, "Hmmm?", while looking down between the boxes with a confused and disappointed facial expression. If the infant did not respond by either touching or unlocking a box, E1 used a set of prompts in order to encourage the infant to help E2.

This task was coded as pass or fail, in which a pass required the infant to open the box (or attempt to open the box) that contained the toy. If the infant opened the box that contained the toy, it was assumed that she had an understanding of E2's false belief about the toy's location, as E1 was not aware that the toy was placed in the other box. Seven participants were excluded from this analysis as they reached for a box before the test trial started ($n = 4$), failed to respond to the task ($n = 2$), and failed to pass the training trials of the task ($n = 1$). A Cohen's Kappa coefficient revealed perfect reliability across two independent coders ($\kappa = 1.00$).

2.2. Results and discussion

The results of the false belief task demonstrate that 15 infants out of 41 (37%) opened the correct, full box. The same pass rate of 37% was obtained when infants' first touch to the correct box was coded. A binomial test revealed that infants' performance was not significantly different from chance (.50) ($p = 0.12$). The present results did not replicate those reported by Buttelmann and colleagues (2009), in which 72% of 18-month-olds opened or touched the correct box. One possibility for this failure to replicate the original study might be the methodological difference in the administration of the task. Our task involved the administration on a table, whereas Buttelmann and colleagues administered the task on the floor at a distance of one meter from the infants, and thus they had to walk to one of the boxes in order to help the experimenter. This small methodological change greatly reduced the attrition rate in the present experiment (28%) in comparison to that of the original study (54%). However, this extra distance may have given infants more time to decide which box to open, preventing an impulsive response (Buttelmann, personal communication, July 2014). Furthermore, the lack of replication is unlikely due to the small sample size ($n = 41$), as it is larger than the sample included in the false belief condition in the original study ($n = 25$).

3. Experiment 2

Experiment 2 was conducted as a follow-up study in order to address the possible reasons infants' performance in Experiment 1 did not match that reported in the original study. First, in order to address the low statistical power of Experiment 1, a much larger sample ($n = 97$) was tested. Second, although the task was still administered at a table, the boxes were positioned at the far end of the table, out of reach of the infant. This modification was made in order to allow infants more time to process the experimenter's actions and minimize impulsive responses. In addition, a true belief condition was added to the experiment in order to compare performance across conditions, as in the original study.

3.1. Method

3.1.1. Participants

For the false belief condition, a total of 106 18-month-old infants were tested. Nine participants were excluded due to fussiness ($n = 5$), parental interference ($n = 2$), inattentiveness ($n = 1$), and failure to respond to the task ($n = 1$), revealing a much lower attrition rate (8%) in comparison to 54% in the original study. The analyses were conducted on a final sample of 97 infants (*Mage* = 18.47 months, $SD = 0.51$; range = 17.4–20; 52 males, 45 females) for the false belief condition. We tested an additional 38 participants in a true belief condition. Five participants were excluded due to failure to respond to the task ($n = 1$) and experimenter error ($n = 4$). The analyses were conducted on a final sample of 33 infants (*Mage* = 18.36 months, $SD = 0.84$; range = 17-20.30; 17 males, 16 females). Participants were exposed to English or French, and were free of auditory or visual impairment.

3.1.2. Materials and procedure

The materials and procedure for the false belief condition were identical to those of Experiment 1, with the exception of the position of the boxes (approx. 50 cm away from the child). Additionally, once E2 tried to open the box and displayed disappointment and confusion, E1 pushed the boxes closer to the infant (approx. 15 cm away from the child). As in Experiment 1, the task was administered at the table to avoid changes in the set up since there were other tasks administered as part of a larger study. No order effect was observed. With regard to the true belief condition, the only change from the procedure of the false belief condition is that E2 remained in the room while E1 switched the location of the toy to the other box. When switching the toy, E1 alternated her eye gaze between the box, the child, and E2, and E2 said "Ahaa!" at each step of the switch to make sure that the infant would notice that she was attentive. E2 then looked away (tied her shoes) while E1 locked the boxes. Following this, E2 went to close the door in order to be in the same position as in the false belief procedure. A Cohen's Kappa coefficient revealed perfect reliability across two independent coders for both conditions ($\kappa = 1.00$).

3.2. Results and discussion

The results of the false belief task demonstrate that 56 out of 97 infants (58%) opened the correct box. A binomial test revealed that infants' performance was not significantly different from chance (.50) ($p = 0.16$). Infants' first touch to a box was also coded. An additional seven participants were excluded from these analyses as they demonstrated an ambiguous response by touching both boxes. By using a touch criterion, results indicate that 50 infants out of 90 infants (56%) touched the correct box, a value not significantly different from chance ($p = 0.29$). With regard to the true belief task, 11 out of the 33 infants (33%) opened the correct, empty box. A binomial test revealed that infants' performance was below chance at the trend level (.50) ($p = 0.08$). Infants' first touch to a box was also coded. A final sample of 34 participants was included in the analyses using a touch criterion, as two participants were excluded because they demonstrated an ambiguous response by touching both boxes. By using a touch criterion, results indicate that 13 out of 34 infants (38%) touched the correct box, a value that was not significantly different from chance ($p = 0.23$).

Furthermore, the box that infants opened was not significantly different across conditions, $\chi^2(1) = 0.82$, p = 0.37 (see Fig. 1). In other words, infants opened the box where the toy was currently located significantly more often than the empty box in both the true belief and false belief conditions. Similar results were found when examining the box that infants first touched, $\chi^2(1) = 0.39, p = 0.53$.
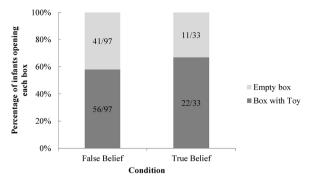
Fig. 1. Percentage of infants who opened each box across the true belief and false belief condition in Experiment 2.

In summary, the results of Experiment 2 revealed that even with a very large sample size, infants' performance on the false belief task is inconsistent with the performance reported in the original study by Buttelmann and colleagues (2009). Furthermore, although allowing infants slightly more time to process which box to open was expected to produce a higher pass rate compared to Experiment 1, performance was still not above chance. However, it is important to note that the pass rate of Experiment 2 was significantly higher than the pass rate of Experiment 1, $U = 1568$, $p = 0.02$. Additionally, infants' performance on the true belief task was inconsistent with the performance observed in the original study, with a much lower pass rate. When comparing the true belief and false belief task, results were also contrary to Buttelmann and colleagues' (2009) findings, where 84% of infants helped the experimenter open the empty box in the true belief condition whereas 72% of infants went to the box where the toy was currently located in the false belief condition. The present results conflict with those of Priewasser and colleagues' (2017), who reported a significant difference in the box that infants opened across the true belief and false belief conditions. However, the current findings are consistent with their results in the true belief condition, as infants' performance on the true belief task in both studies is at chance.

4. General discussion

It is well established that scientific progress requires the replication of theoretically important findings. A recent open collaboration systematically examined the reproducibility of a sample of 100 published studies. Their findings demonstrate that a large proportion of replications produced weaker effects in comparison to the original findings (Open Science Collaboration et al., 2015). With regard to false belief understanding in infancy, there have been few systematic attempts to replicate the results from studies using common false belief tasks. In addition to this issue, most of the published research comes from a single or a handful of laboratories, increasing the need for replication. To our knowledge, only two replication studies attempting to examine false belief in infants using a helping task have been conducted and have yielded mixed results (Powell, Hobbs, Bardis, & Carey, 2017; Priewasser et al., 2017). Unfortunately, these studies often included a small sample size.

The main objective of the present study was to attempt a conceptual replication of the false belief task based on the prompted-action, helping paradigm. The replication of studies which are based on the mentalistic view is a critical step in addressing the current debate regarding the depth of infants' theory of mind (Poulin-Dubois & Yott, 2017). According to one view, infants have a comparable understanding of false belief to preschoolers or adults (Carruthers, 2013; Baillargeon et al., 2016), but task demands (verbal or executive function skills) mask their competence. The alternate view is that infants do not have a mentalistic perspective of false belief as their behavior can be explained by a low-level analysis of the display or by simple behavioral rules (Apperly & Butterfill, 2009; Heyes, 2014; Perner & Ruffman, 2005; Ruffman, 2014; Ruffman & Perner, 2005; Ruffman, Taumoepeau & Perkins, 2012). Despite the inclusion of a large sample size, particularly in Experiment 2, we failed to replicate previous findings demonstrating that a majority of 18-month-olds pass the false and true belief task. Thus, what appears to be a slight methodological change (distance between the boxes and the child) significantly impacted infants' performance, suggesting that performance on this false belief helping task is not robust. The impact of this methodological change on the results suggests that infants might perform without considering the experimenter's false or true belief.

Although the results of our true belief condition do not replicate findings from the original study, Priewasser and colleagues (2017) also did not replicate the true belief condition, as infants' performance on this task was not significantly different from chance. Infants' performance was also not significantly different from chance when a third box was introduced in their new true belief condition. Importantly, with the new true and false belief conditions, Priewasser and colleagues (2017) challenged Buttelmann and colleagues' mentalistic interpretation, and found support for a teleological interpretation. Specifically, they argue that infants simply inferred the experimenter's interest in the toy, and subsequently helped her find it when she looked in the third box, without the need to attribute any false belief to the experimenter.

In addition, a factor that might have impacted performance on the helping task is individual differences in social skills. According to Allen (2015), children may be relying on the broader social context, such as playing a trick on the experimenter, when choosing which box to open, instead of reasoning about the experimenter's false belief. Using two control conditions, Allen provided evidence against the mentalistic interpretation by demonstrating that children's performance was influenced by the nature of the social context. Unfortunately, our design did not include any measure of infants' sensitivity to trickery, and therefore, this "lean" alternative

interpretation to Buttelmann and colleagues' findings remains to be directly tested. Nonetheless, the current results seem to challenge Allen's (2015) findings, as infants in the present study were not significantly more likely to touch and/or open the box containing the toy. It is worth noting that in order to rule out the role of "sneakiness" or "trickery", Buttelmann and colleagues (2014) designed a procedure that did not involve trickery through the use of an unexpected-content task, instead of a location false belief task. The authors reported that approximately 66% of the infants passed the task in the false belief condition, although this was a performance still not significantly different from chance. Thus, although helping tasks present advantages over tasks based on visual responses (i.e., simple reactions to novelty are ruled out), they pose unique challenges that seem to make performance unstable.

Similarly, individual differences in prosocial skills might have influenced infants' performance on this task. For example, the inclination to help the experimenter find her toy based on her false belief requires a willingness to help a stranger who expresses disappointment, which is a prosocial skill that gradually develops during the second year of life (Svetlova, Nichols, & Brownell, 2010). Although infants' performance on the true and false belief task was inferior to that reported in the original study, infants' willingness to help the experimenter was improved when the task required infants to extend their arm in order to reach the target box instead of walking towards the boxes. This yielded a considerably lower attrition rate than the original study by Buttelmann and colleagues (2009) (i.e., Experiment 2: 8% vs. Buttelmann et al.: 54%).

The present findings also contribute to the debate regarding the nature of helping in infancy. According to Paulus (2014), there are four classes of models that help explain the development of prosocial behavior in young children. The first model proposed is emotion-sharing, which assumes that children engage in prosocial behavior due to self-other differentiation and their ability to regulate negative emotions and engage in solution-oriented behaviors (i.e., comforting). The second model proposed is goal-alignment, which assumes that infants take over the other individual's goal as if it were theirs, which motivates the infants to act prosocially. The third model proposed is social interaction, which assumes that infants act prosocially because they enjoy interacting with others, and not because they want to genuinely help others. Lastly, the fourth model proposed is social-normative, which assumes that infants' environment fosters and supports the development of prosocial behavior. The helping task used in the current study is probably one of the most demanding instrumental helping tasks. An instrumental helping task usually involves assisting another individual in achieving an action-based goal such as obtaining an object that is out of reach, which appears by 12–14 months of age (Liszkowski, Carpenter, Striano, & Tomasello, 2006; Warneken & Tomasello, 2007). There may be more cognitive demands in the current task than in the original one, as the goal is no longer visible and the infant has to maintain in working memory what the protagonist wants, that is, find her toy. It is well known that instrumental helping tasks vary in difficulty. For instance, infants are more inclined to help during tasks where there is an object out of reach compared to tasks involving a physical obstacle, as the goal is easier for infants to comprehend when the experimenter tries to reach for an object (Warneken & Tomasello, 2006). According to Paulus (2014), infants' instrumental helping may be motivated by goal-alignment, which suggests that goal understanding is sufficient for infants to act prosocially. Therefore, according to this view, infants' helping behavior may not be motivated by understanding others' mental states (Paulus, 2014). This lean interpretation appears to be supported by the findings of a recent study in which performance on the current false belief task was unrelated to performance on the false belief task based on the violation of expectation paradigm (Poulin-Dubois & Yott, 2017).

Aside from individual differences in prosocial skills, there could be many reasons for a lack of replication (Gilbert, King, Pettigrew, & Wilson, 2016). For instance, testing a population that is different from the one included in the original sample may significantly affect the results. Specifically, the current study tested Canadian infants, whereas the original study examined the performance of German infants and children. Although language and/or culture have been shown to have an impact on the development of theory of mind (e.g., Shahaeian, Peterson, Slaughter, & Wellman, 2011), such effects have only been observed in preschool children, who have more advanced verbal skills. Given that both language groups had minimal verbal skills, language is unlikely to be the reason for our non-replication. Another established reason leading to a lack of replication is the use of a procedure that differs from the original study's procedures in substantial ways (Open Science Collaboration et al., 2015). In the present study, the experimental set-up differed from the original one, in that the task was administered at a table instead of on the floor, due to low compliance observed both in the original study and in our pilot testing. What seemed to be a minor change in the procedure may account for our failure to replicate. Future replications should carefully ensure that the methodology and procedure are identical to that of Buttelmann and colleagues' (2009) study by administering the task on the floor, even if it requires that infants move toward the display and of the high attrition rate that ensues. Finally, other tasks (i.e., a task examining social skills in infants) should also be administered to better understand the source of individual variability. Notwithstanding these issues, the fact that performance dropped with such a minor change in methodology suggests that if false belief is mastered in infancy, it must be a rather fragile concept in comparison to older children and adults.

References

Allen, J. W. P. (2015). How to help: can more active behavioral measures help transcend the infant false belief debate? *New Ideas in Psychology, 39*, 63–72. http://dx. doi.org/10.1016/j.newideapsych.2015.07.008.

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

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110–118. http://dx.doi.org/10.1016/j.tics.2009. 12.006.

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology, 67*, 159–186. http://dx.doi.org/10.1146/annurev-psych-010213-115033.

Burnside, K., Ruel, A., Azar, N., & Poulin-Dubois, D. (2017). Implicit false belief across the lifespan: Non-replication of an anticipatory looking task. *Cognitive Development*. http://dx.doi.org/10.1016/j.cogdev.2017.08.006 Advanced Online Publication.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112*(2), 337–342. http://dx.doi.org/10.1016/j.cognition.2009.05.006.

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

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language, 28*(2), 141–172. http://dx.doi.org/10.1111/mila.12014.

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2017). *Does the infant really have a ToM? Replicability and validity of 4 implicit measures.* Vienna, Austria: Poster session presented at the International Convention of Psychological Science.

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. (2016). Comment on "Estimating the reproducibility of psychological science". *Science, 351*(6277), 1037. http://dx. doi.org/10.1126/science.aad7243.

Heyes, C. (2014). False belief in infancy: a fresh look. *Developmental Science, 17*(5), 647–659. http://dx.doi.org/10.1111/desc.12148.

Knudsen, B., & Liszkowski, U. (2012a). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy, 17*(6), 672–691. http://dx.doi.org/10.1111/j.1532-7078.2011.00105.x.

Knudsen, B., & Liszkowski, U. (2012b). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science, 15*(1), 113–122. http://dx.doi.org/10.1111/j.1467-7687.2011.01098.x.

Kovács, A. M., Téglás, E., & Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830–1834. http://dx.doi. org/10.1126/science.1190792.

Kulke, L., & Rakoczy, H. (2017). How reliable and valid are anticipatory looking measures in theory of mind task? In H. Rakoczy & L. Kulke (Chairs), *Are implicit theory of mind findings robust? Some doubts from converging non-replications across the lifespan*. Symposium conducted at the Society for Research in Child Development Biennial Meeting. Austin, Texas.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12-and 18-month-olds point to provide information for others. *Journal of Cognition and Development, 7*(2), 173–187. http://dx.doi.org/10.1207/s15327647jcd0702_2.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255–258. http://dx.doi.org/10.1126/science.1107621.

Open Science Collaboration, Nosek, B. A., Aarts, A. A., Anderson, C. J., Anderson, J. E., ... Kappes, H. B. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. http://dx.doi.org/10.1126/science.aac4716.

Paulus, M. (2014). The emergence of prosocial behavior: why do infants and toddlers help, comfort, and share? *Child Development Perspectives, 8*(2), 77–81. http://dx. doi.org/10.1111/cdep.12066.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: how deep? *Science, 308*(5719), 214–216. http://dx.doi.org/10.1126/science.1111656.

Perner, J. (2014). Commentary on Ted Ruffman's 'To belief or not belief …'. *Developmental Review, 34*, 294–299. http://dx.doi.org/10.1016/j.dr.2014.05.002.

Poulin-Dubois, D., & Yott, J. (2017). Probing the depth of infants' theory of mind: disunity in performance across paradigms. *Developmental Science;* e12600. http://dx. doi.org/10.1111/desc.12600.

Poulin-Dubois, D., Polonia, A., & Yott, J. (2013). Is false belief skin-deep? The agent's eye status influences infants' reasoning in belief-inducing situations. *Journal of Cognition and Development, 14*(1), 87–99. http://dx.doi.org/10.1080/15248372.2011.608198.

Powell, L., Hobbs, K., Bardis, A., & Carey, S. (2017). Investigating the reproducibility of implicit theory of mind tasks with different levels of representational demand. In H. Rakoczy & L. Kulke (Chairs). *Are implicit theory of mind findings robust? Some doubts from converging non-replications across the lifespan.* Symposium conducted at the Society for Research in Child Development Biennial Meeting, Austin, Texas.

Priewasser, B., Rafetseder, E., Gargitter, C., & Perner, J. (2017). Helping as an early indicator of a theory of mind: mentalism or teleology? *Cognitive Development. Advanced Online Publication.* http://dx.doi.org/10.1016/j.cogdev.2017.08.002.

Ruffman, T., & Perner, J. (2005). Do infants really understand false belief? Response to Leslie. *Trends In Cognitive Sciences, 9*(10), 462–463. http://dx.doi.org/10.1016/ j.tics.2005.08.001.

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology, 30*(1), 87–104. http://dx.doi.org/10.1111/j.2044-835X.2011.02045.x.

Ruffman, T. (2014). To belief or not belief: children's theory of mind. *Developmental Review, 34*(3), 265–293. http://dx.doi.org/10.1016/j.dr.2014.04.001.

Schneider, D., Slaughter, V. P., & Dux, P. E. (2015). What do we know about implicit false-belief tracking? *Psychonomic Bulletin & Review, 22*(1), 1–12. http://dx.doi. org/10.3758/s13423-014-0644-z.

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology, 47*(5), 1239–1247. http://dx.doi.org/10.1037/a0023899.

Sodian, B., Licata, M., Kristen-Antonow, S., Paulus, M., Killen, M., & Woodward, A. (2016). Understanding of goals, beliefs, and desires predicts morally relevant theory of mind: a longitudinal investigation. *Child Development, 87*, 1221–1232. http://dx.doi.org/10.1111/cdev.12533.

Sodian, B. (2011). Theory of mind in infancy. *Child Development Perspectives, 5*(1), 39–43. http://dx.doi.org/10.1111/j.1750-8606.2010.00152.x.

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

Svetlova, M., Nichols, S. R., & Brownell, C. A. (2010). Toddlers' prosocial behavior: from instrumental to empathic to altruistic helping. *Child Development, 81*(6), 1814–1827. http://dx.doi.org/10.1111/j.1467-8624.2010.01512.x.

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*(1), 172–187. http://dx.doi.org/10.1111/j.2044-835X.2011.02067.x.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science, 311*(5765), 1301–1303. http://dx.doi.org/10.1126/ science.1121448.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11*(3), 271–294. http://dx.doi.org/10.1111/j.1532-7078.2007. tb00227.x.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development, 72*(3), 655–684. http:// dx.doi.org/10.1111/1467-8624.00304.

Wellman, H. M. (2014). Making minds: how theory of mind develops. *Oxford University Press.*. http://dx.doi.org/10.1093/acprof:oso/9780199334919.001.0001.

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*(5), 683–698. http://dx.doi.org/10.1080/15248372.2015.1086771.