PAPER

WILEY **Developmental Science**

Probing the depth of infants' theory of mind: disunity in performance across paradigms

Diane Poulin-Dubois | Jessica Yott

Department of Psychology, Concordia University, Montréal, Québec, Canada

Correspondence
Diane Poulin-Dubois, Department of Psychology, Concordia University, 7141 Sherbrooke St. West, PY-170, Montréal, Québec H4B 1R6, Canada.
Email: diane.poulindubois@concordia.ca

Abstract

There is currently a hot debate in the literature regarding whether or not infants have a true theory of mind (ToM) understanding. According to the mentalistic view, infants possess the same false belief understanding that older children have but their competence is masked by task demands. On the other hand, others have proposed that pre-verbal infants are incapable of mental state attribution and simply respond to superficial features of the events in spontaneous-responses tasks. In the current study, we aimed to clarify the nature of infants' performance in tasks designed to assess implicit theory of mind (ToM) by adopting a within-subject design that involved testing 18-month-old infants on two batteries of tasks measuring the same four ToM constructs (intention, desire, true belief, and false belief). One battery included tasks based on the violation-of- expectation (VOE) procedure, whereas the other set of tasks was based on the interactive, helping procedure. Replication of the original findings varied across tasks, due to methodological changes and the use of a within-subject design. Convergent validity was examined by comparing performance on VOE and interactive tasks that are considered to be measures of the same theory of mind concept. The results revealed no significant relations between performance on the pairs of tasks for any of the four ToM constructs measured. This pattern of results is discussed in terms of current conflicting accounts of infants' performance on implicit ToM tasks. A video abstract of this article can be viewed at: https://www.youtube.com/watch?v=U3vqfe_zdhA&feature=youtu.be

RESEARCH HIGHLIGHTS

- Infants aged 18 months were administered two batteries of theory of mind tasks based on either the violation-of-expectation (VOE) or the helping paradigms.
- The performance on some of the original VOE and helping tasks was not replicated, most likely due to the within-subject design.
- Results indicate no link in infants' performance across paradigms, showing a lack of convergent validity among tasks measuring similar theory of mind concepts in infancy.

1 | INTRODUCTION

The ability to explain and predict other people's behavior by understanding others' mental representation of the world is a core system in human social cognition. The development of this ability, also known as a Theory of Mind (ToM), has been well researched over the past 30 years (Wellman, 2014). Although a theory of mind entails being able to reason about a number of mental states (e.g., intention, emotion, desire) it is the false belief task, which examines children's expectation about how people behave when they have representations that are inconsistent with reality, which has been considered the litmus test of theory of mind. Decades of research converged on a developmental pattern: below the age of 3 years children reliably fail the false belief task (expecting people to behave according to reality) and reliably pass this task by 5 years of age (expecting people to behave according to their false beliefs) (Wellman, Cross, & Watson, 2001).

Although there is a consensus that children pass the traditional, explicit false belief task during the preschool period, research over the past decade suggests that even infants might possess some form of understanding of people's mind (see Poulin-Dubois, Brooker, &

Chow, 2009 and Sodian, 2011, for reviews). This idea originated with Clements and Perner's (1994) demonstration that although 3-year-old children fail the traditional verbal false belief task, they seem to know the correct answer when their non-verbal behaviors are monitored (e.g., eye gaze). This striking finding triggered studies investigating the development of ToM understanding in infancy using implicit measures, such as the Violation of Expectation (VoE) paradigm. The first studies conducted with infants showed that goal and desire understanding can be documented between 6 and 18 months of age (Poulin-Dubois, 1999; Woodward, 1998). This paradigm was successfully adapted to test false belief by Onishi and Baillargeon (2005) who demonstrated that 15-month-old infants looked significantly longer at an incongruent test trial, where an actor's actions were inconsistent with her false belief (searching for an object at the current location), than at a congruent trial (searching at the previous location). These results have been replicated and extended to children as young as 7 months of age using VOE or anticipatory looking procedures (He, Bolz, & Baillargeon, 2011; Kovács, Téglás, & Endress, 2010; Phillips et al., 2015; Scott, Baillargeon, Song, & Leslie, 2010; Senju, Southgate, Snape, Leonard, & Csibra, 2011; Song, Onishi, Baillargeon, & Fisher, 2008; Southgate, Senju, & Csibra, 2007; Surian, Caldi, & Sperber, 2007; Surian & Geraci, 2012; Yott & Poulin-Dubois, 2012). This body of research has provided the foundation for the provocative claim that infants possess an abstract capacity to reason about false beliefs (Baillargeon, Scott, & He, 2010; Scott, 2017; Scott, Richman, & Baillargeon, 2015).

It has been proposed that infants perform better on the violation-of-expectation task and other tasks based on spontaneous responses because, unlike the traditional elicited-response task, these tasks involve less heavy processing demands (Baillargeon et al., 2010; Scott, 2017). More specifically, in order to answer the question correctly in the Sally/Ann task, children face three executive function demands: response-selection, response-inhibition, and working memory. Although non-elicited-response tasks might involve some of these demands, it is argued that they impose fewer demands on children, allowing them to express their false belief understanding. Thus, according to this mentalistic view, infants possess the same false belief understanding that older children have but their competence is masked by task demands. This rich interpretation of non-elicited responses in implicit ToM tasks is currently the subject of a hot debate (Apperly & Butterfill, 2009; Heyes, 2014; Perner, 2010; Ruffman, 2014; Ruffman & Perner, 2005). Some researchers have proposed that infants' behaviors in these tasks are explained by basic cognitive abilities. Among these leaner arguments is that infants' looking time patterns in these implicit tasks can be explained by the violation of learned behavioral rules (Ruffman & Perner, 2005). For example, infants might look longer in the incongruent trial of the false belief task because it violates the well-learned rule that people look for objects in the last place they saw them. Others have proposed that infants' looking time patterns in implicit paradigms simply reflect sensitivity to low-level perceptual or attentional factors, and not to an actor's true or false belief about an object's location (Heyes, 2014; Phillips et al., 2015). For example, it has been argued that low-level changes in the properties of the stimuli encoded by the child during the familiarization and belief induction trials

(e.g., colors, movements, and shapes) could account for longer looking times during the test trials (Heyes, 2014; but see Scott & Baillargeon, 2014). A third alternative interpretation of infants' behaviors in false belief tasks has proposed a dissociation between a minimalist theory of mind that allows for infants (and adults) to track belief-like states or registrations of an agent and a full-fledged ToM that represents mental states such as false beliefs (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). The first system is developmentally precocious, fast, automatic, and relatively independent of central cognitive resources. In contrast, the second system develops late, operates slowly, and recruits executive control processes. The difference in representational capacities of these two "mindreading" systems is reflected in differential patterns of performance in a number of tasks (see Low, Apperly, Butterfill, & Rakoczy, 2016, for a review). For instance, a dissociation has been reported between children's performance across two tasks of perspective-taking problems: level-1 (judging whether someone sees an object) and level-2 (judging how someone sees an object) (Surtees, Butterfill, & Apperly, 2012).

Recently, prompted-action tasks based on helping behaviors have been developed to test infants' false belief understanding with the aim to address the criticisms toward the tasks based on looking time patterns. The typical paradigm involves prompting the infant to select or retrieve an object for the agent, with success requiring the infant to understand the agent's false belief about the location or identity of the object. For example, the experimenter first shows the infant how to unlock two boxes, and then an agent enters the room and hides a toy in one of the two boxes before leaving. In her absence, the experimenter moves the toy to the other box and then locks both boxes. When the agent returns and tries unsuccessfully to open the locked box where she believes the toy is located, 18-month-old infants offer help by opening the correct box (Buttelmann, Carpenter, & Tomasello, 2009). Similar helping behaviors have been reported in referential communication tasks when the agent incorrectly points at a box where she believes her toy is located (Southgate, Chevallier, & Csibra, 2010). In anticipatory-pointing tasks, infants point to inform an agent that aversive material has been placed at the location that the agent falsely believes her desired toy is (Knudsen & Liszkowski, 2012a, 2012b). As in research based on anticipatory looking or VOE procedure, these findings have been interpreted as reflecting a rich understanding of the mind. However, this mentalistic interpretation of infants' behaviors has also been challenged as reflecting a simple understanding of the social context (Allen, 2015).

In summary, it has been argued that findings from studies based on spontaneous visual responses and prompted-action tasks support the view that early psychological reasoning is mentalistic, that is, infants possess an abstract understanding of mental states, including motivational, epistemic, and counterfactual states (Baillargeon, Scott, & Bian, 2016). As previously discussed, this rich interpretation has been challenged with a number of alternative interpretations. There is also a body of research, albeit limited, which provides some evidence that it may be premature to conclude that infants are naïve psychologists. One approach has been to examine the impact of "noise" in the display or the effect of priming events on looking patterns in the

VOE paradigm. For example, regarding the behavioral rule account, Yott and Poulin-Dubois (2012) taught 18-month-olds a new rule: objects are never at the last place you saw them. When subsequently tested on the VOE version of the false belief task, infants displayed the same pattern of looking as in the original study, providing no support for this alternative proposal. However, the authors concluded that a few training trials might not have been strong enough to overwrite a well-learned rule. Other attempts have also been made to test the depth of false belief reasoning in infancy. In a recent study, 14- and 18-month-old infants' reasoning about false belief was tested with the violation-of-expectation paradigm, except that transparent boxes replaced opaque boxes, making the location of the object visible to both the infant and agent before and after the object location change (Poulin-Dubois, Polonia, & Yott, 2013). Infants looked longer when the agent searched in the empty box, the opposite looking pattern than in the original task with opaque boxes, indicating that they updated their prediction of the agent's actions as a function of her current visual perspective. In a second experiment, the original false belief violation-of-expectation procedure was used, except that the agent wore a blindfold during the induction and test trials. This methodological change, which should not impact the attribution of false belief, made infants look equally long at the congruent and incongruent test trials, challenging the view that infants possess a mature concept of false belief. Another approach for understanding the depth of infants' ToM has been to conduct longitudinal studies to assess continuity in implicit and explicit measures of ToM. For example, infants' social information processing (decrement of attention in habituation tasks measuring goal-encoding) was found to predict children's performance on a standard ToM battery many years later (Aschersleben, Hofer, & Jovanovic, 2008; Wellman, Lopez-Duran, LaBounty, & Hamilton, 2008; Wellman, Phillips, Dunphy-Lelii, & LaLonde, 2004; Yamaguchi, Kuhlmeier, Wynn, & van Marle, 2009). Recently, a long-term longitudinal study of ToM development assessed implicit false belief understanding with the anticipatory looking measure at the age of 18 months and found a correlation with explicit false belief understanding at 48 months (Thoermer, Sodian, Vuori, Perst, & Kristen, 2012). At first sight, this finding suggests shared variance between infant and preschool false belief assessments and appears to provide support for a conceptual interpretation of infants' ToM competencies. However, the link was limited to a specific false belief task and the observed continuity might simply reveal one of the socio-cognitive precursors to explicit false belief, similar to the link between early joint attention and later language skills (Brink, Lane, & Wellman, 2015).

Finally, an adult-like ToM implies that coherence in reasoning about different mental states should be observed, as is the case for explicit theory of mind (Gopnik & Astington, 1988; Rakoczy, Bergfeld, Schwarz, & Fizke, 2014, but see Rice & Redcay, 2015). A comparison of performance across a wide range of standard ToM constructs has revealed moderate to strong inter-task relations in preschool children (Carlson, Mandell, & Williams, 2004; Hughes & Ensor, 2005). In toddlers, Chiarella, Kristen, Poulin-Dubois, and Sodian (2013) reported no significant correlations among scores on ToM tasks measuring different constructs in 30- to 38-month-olds. More specifically, children completed two or three ToM tasks, including a visual perspective-taking task, a desire-understanding task, and an emotional perspective-taking task, and results revealed no significant correlations in either a Canadian or a German sample. More relevant is a recent study that examined inter-task comparisons in infancy with all tasks based on the VOE paradigm (Yott & Poulin-Dubois, 2016). Using a within-subject design, infants aged 14 and 18 months were administered tasks measuring intention, true belief, desire, and false belief understanding over two testing sessions. Inter-task comparisons revealed only a significant correlation between looking time at the false belief and intention tasks. Although these results appear to weaken the conceptual view about infants' ToM, it could be argued that a lack of integration is inconclusive since poor integration has sometimes been reported at older ages.

In the current study, we adopted a new approach to address the current debate regarding the nature of the competence that infants possess about the mind. We addressed this issue by adopting a within-subject design that involved testing infants on two batteries of tasks measuring the same implicit theory of mind constructs, with one battery based on the VOE procedure and the other on the spontaneous helping procedure. To our knowledge, only a few studies have compared infants' performance on ToM abilities measured with different paradigms and all these studies have used a prospective research design (Olineck & Poulin-Dubois, 2009; Thoermer et al., 2012). We reasoned that if infants' performance of these tasks can be accounted for by a fully representational theory of mind, then one would expect concurrent convergent validity among tasks measuring the same mental state understanding, regardless of the type of measurement. Of course, given that processing demands vary across tasks designed to test early theory of mind, one would expect that group performance would vary across tasks. This is recognized in the recent processing-load account of infants' false belief reasoning. According to this revised mentalistic view, the capacity to represent false beliefs emerges in infancy as indicated by infants' success on spontaneous-response tasks. However, this precocious competence might be masked in a given false belief task if the processing demands of the task exceed the child's processing abilities (Scott, 2017; Scott & Roby, 2015). Despite expected variability across tasks, what is more critical is the predicted *stability* of performance when individual results are considered, with an individual's relative performance preserved across paradigms if a mental state is attributed to the agent. In contrast, if infants represent the test events shown in the violation of expectation paradigm as new spatiotemporal relations among colors, shapes, and movements (perceptual novelty) and not as actions on objects by agents, no systematic relation should be observed between the scores on the VOE tasks and those based on prompted-action tasks (Heyes, 2014). With regard to the account based on learning of behavioral rules, it is assumed to provide an explanation for data generated by both paradigms because this proposal posits that infants succeed on all ToM tasks through a combination of pre-experimental experience of people's actions and observations in the familiarization phase (Ruffman, 2014). However, in spontaneous helping tasks, there are no familiarization phases similar to those in VOE tasks, so that pre-experimental experience is critical

for infants' interpretation of the agent's behaviors. Consequently, poor convergent validity would also be predicted by such an account. Finally, it is our understanding that the minimalist, two-systems perspective would predict poor coherence across tasks based on spontaneous responses versus prompted actions given that the early-developing system involves rapid online mindreading triggered by direct cues, such as line of sight (Low et al., 2016).

2 | METHOD

2.1 | Participants

The sample consisted of 53 18-month-old infants (32 males), with a mean age of 1;6 (range = 1;4 to 1;8). The sample was ethnically diverse and on the basis of parental reports, infants had no visual or auditory impairments, and had a minimum 35-week gestational period. All infants were recruited from birth records provided by a governmental health services agency and were exposed to primarily English or French. The current sample is part of a larger sample included in a recently published study on the convergence among a range of VOE constructs (e.g., intention vs. desire) in 14- and 18-month-old infants (Yott & Poulin-Dubois, 2016). Only the 18-month-old infants who completed both helping and VOE tasks for each of the ToM constructs were included in the present sample.

An additional 19 (26%) 18-month-old infants participated but were excluded from the study due to fussiness ($n = 8$), parental interference ($n = 2$), technical difficulties ($n = 2$), a reported developmental delay ($n = 1$), and missing one of the two testing sessions ($n = 6$).

2.2 | Materials

2.2.1 | VOE tasks

A stage-like apparatus (107 cm × 61 cm × 211 cm) was used to administer all four tasks.

Intention task

A black barrier (30 cm × 25 cm) was used and placed on the right side of the stage. At the beginning of the task, a small yellow duck (12 cm × 12 cm) was placed on the right side of the barrier. The experimenter could not see the duck as it was out of sight and hidden by the barrier.

Belief tasks

A red cup (7.5 cm diameter, 10.5 cm height) or a yellow duck (11 cm × 11 cm) was placed on the surface top directly between a yellow and a green box (14 cm × 14 cm × 14 cm). The distance between the boxes was 18 cm. Each box had an opening on the side facing the cup or duck (14 cm × 14 cm) that was covered with a fabric fringe. A rectangular opening underneath each box allowed for the attraction between a magnet located inside the cup and duck (2.5 cm × 5 cm length × 0.6 cm) and a magnet under the stage, operated by the experimenter (7.6 cm diameter). The magnet was used to unobtrusively transfer the cup or duck from one box to the other underneath the stage.

Desire task

Two food pairings were used during the desire task: lettuce and Honey Nut Cheerios, and broccoli with Pepperidge Farm Gold Fish crackers. Both food pairings were placed in their own clear plastic containers.

2.2.2 | Interactive tasks

During all tasks, the child was seated across the table from the experimenter.

Intention task

The materials consisted of five novel object pairs. Each pair of objects could be used to complete an intended target action demonstrated by the experimenter.

Belief tasks

Two wooden boxes with lids were used for this task. Both boxes could be locked using wooden pins, and had handles on the lid. A plastic toy caterpillar was used as the toy being hidden in the boxes.

Desire task

During this task, the participants were presented with two sets of clear plastic containers holding food. This first set consisted of Cheerios cereal and lettuce, and the second set consisted of Pepperidge Farm Goldfish crackers and broccoli.

2.3 | Design and procedure

Infants and their parents were invited to the laboratory for two testing sessions that each lasted approximately 45 minutes, scheduled one to two weeks apart. The child was seated across the table from the experimenter during the interactive tasks and directly in front of the stage-like apparatus for the VOE tasks. During all tasks, the child was seated either in a high chair, or on the parent's lap. On the first visit, parents were brought to a reception room, and were asked to complete a consent form, a demographic questionnaire, and the receptive vocabulary checklist, while the child familiarized with the experimenters and the environment. The Level-1 short form of the McArthur-Bates Communicative Inventory was administered in French or English to control for infants' verbal ability (Fenson et al., 2000; Trudeau, Frank, & Poulin-Dubois, 1999). All tasks were recorded in order to code infants' responses off-line, as well as code inter-rater reliability. All families were offered $20 in financial compensation per session for their participation in this study.

All infants participated in four VOE tasks and four interactive tasks. To prevent boredom and carry-over effects, half of the interactive and VOE tasks were administered in each session with the congruent and incongruent trials of the VOE tasks always administered in separate sessions. The VOE and interactive tasks were administered in a block, but the order of the tasks/trials was counterbalanced, creating eight different orders. No order effects were observed.

2.3.1 | VOE tasks

Intention task

The intention task was adapted from Phillips and Wellman (2005). This task consisted of seven trials, which included three familiarization trials and four test trials. During the first familiarization trial, a black barrier separated the experimenter from a yellow duck. Each familiarization trial began with an attention-attracting noise and the curtain rising. During the demonstration phase, which lasted approximately 2 seconds, the experimenter reached over the barrier, grabbed the duck, and held it in front of her while gazing at it. Trials were coded live and began once the experimenter paused with the duck in her hands. The trial lasted a maximum of 30 seconds, but ended if the infant looked away from the display for more than 2 consecutive seconds after looking at the display for a minimum of 2 cumulative seconds, or if he/she looked away for 10 consecutive seconds. During the test trials, the same procedure was used except that the barrier was removed from the stage. Of the four test trials, two trials were congruent and two were incongruent. During the congruent trials, the experimenter reached directly for the duck and then held it in front of her. In contrast, during the incongruent trial, the experimenter reached for the duck indirectly by following the same path as though the barrier was present. This reach was considered incongruent because the experimenter no longer needed to follow this path.

Belief tasks

Infants participated in a Full Box belief task and an Empty Box belief task, each completed on a separate day. The belief tasks were adapted from Onishi and Baillargeon (2005) to examine infants' understanding of true and false beliefs. During each of the belief tasks, all infants completed three familiarization trials, followed by a false belief induction trial and a false belief test trial, and then a true belief induction trial followed by a true belief test trial. A curtain was raised and lowered at the beginning and end of each trial and was accompanied by an attention-attracting sound.

During the 8-second familiarization trial, the experimenter raised the curtain, picked up the cup and placed it inside one of the two boxes. Once the cup was hidden, the experimenter paused with her hand inside the box. A trial lasted a maximum of 8 seconds and ended if the infant looked away from the display for more than 2 consecutive seconds after looking at the display for a minimum of 2 cumulative seconds. In addition, if infants looked away for 10 consecutive seconds before having looked at the screen for the minimum 2 seconds, the trial ended. During the second and third familiarization trials, the experimenter reached into the box where the cup was hidden and then paused with her hand inside the box until the trial ended.

During the false belief induction trial, the cup moved from one box to the other through a magnet operated by the experimenter. Next, during the false belief test trial, the experimenter reached into the full box (the box with the cup), then paused. This test trial was incongruent, because the experimenter's behavior (searching in the full box) was incongruent with her belief (no knowledge of the cup's new location). Next, infants observed a true belief induction trial, where the experimenter moved the cup back to the first box using the magnet. In contrast to the false belief induction trial, the experimenter observed the cup move from one box to the other. Lastly, infants observed the true belief test trial during which the experimenter again reached into the full box (the box with the cup). This time, the trial was congruent, because the experimenter's action (searching in the full box) was consistent with her belief (knowledge of the cup location). This belief task is called the Full Box task, as infants observed the experimenter search in the full box during each test trial.

During the Empty Box belief task, infants observed the same experimenter and trials, although in place of the toy cup, a rubber duck was used. In addition, the experimenter searched in the empty box instead of the full box during the true and false belief test trials. Like the Full Box belief task, infants observed both an incongruent and a congruent trial, but this time, the congruent trial occurred during the false belief test trial, because the experimenter searched in congruence with her belief (the toy's original location). In contrast, the incongruent trial corresponded to the true belief test trial, because the experimenter searched the empty box, after having observed the cup change locations.

Using this within-subjects design, all infants had observed an incongruent and a congruent trial following both the true and false belief induction trials. In addition, infants' individual looking times during the incongruent and congruent trials could be compared for both the true belief and false belief test trials. The order in which the infants completed the belief tasks was counterbalanced. In addition, the design allowed for a congruent belief trial to be presented first, and an incongruent trial to be presented second in one belief task, and vice versa in the second belief task.

Desire task

The current VOE task was adapted from Repacholi and Gopnik's (1997) interactive desire task. Infants participated in two desire conditions; a congruent condition and an incongruent condition. During the congruent desire condition, infants observed three familiarization trials, followed by one test trial. The familiarization trials began with an attention-attracting sound and the curtain rising. Two experimenters (E1 and E2) sat side by side at a table in front of the infant. Placed in front of E2 was one plastic container filled with crackers and a second with broccoli. Three pieces of broccoli and three crackers were placed in front of the other experimenter (E1). The trial began with E1 picking up a cracker, showing it to the infant, and then eating it. After putting the cracker in her mouth, she said with a look of content, "Mmm Crackers, Mmm". She then picked up a piece of broccoli, placed it in her mouth and said, "Eww broccoli, Eww" with a look of disgust. All food items were eaten in this manner while E2 watched with a neutral facial expression. These familiarization trials lasted approximately 20 seconds.

Subsequently, E1 turned to E2, looked up at the full containers in front of her, and said, "Can I have some?" with her hands open in

front of her, palms up. E2 then looked at both containers of food, reached for the crackers, turned toward E1 and placed some in her hand. E1 turned toward the infant and then looked down at the food with a neutral facial expression and paused. At this point, the test trial began and both experimenters remained still for 10 seconds. The same procedure was used for the incongruent desire task, except that the crackers and broccoli were replaced with Cheerios and lettuce. In addition, during the familiarization trials, E1 demonstrated a preference for the Cheerios; however, during the test trial, E2 gave her lettuce. In this way, the incongruent desire task differed from the congruent desire task, because E1 received the food for which she did not demonstrate a preference.

2.3.3 | Interactive tasks

Intention task

The procedure used for this task was based on that used in Meltzoff (1995), Bellagamba, Camaioni, and Colonnesi (2006), and Olineck and Poulin-Dubois (2009). More specifically, infants were only tested on the "Demonstration of Intention" condition of the re-enactment task (Meltzoff, 1995). The task consisted of five test trials, each with one novel object pair. Two distinct presentation orders were used across infants. For this task, the Experimenter 1 (E1) presented the object pair to the child and said, "Hi [Child's name]. Watch, I have something to show you." E1 then modeled the intention to perform an action three times. Importantly, the experimenter did not provide verbal or facial expressions during the demonstrations. For the dumbbell object, the experimenter held a wooden cube in each hand and appeared as though she was trying to pull it apart into two halves. The experimenter failed to do so, however, because one of her hands slipped off the end as she tried to pull. The hand that slipped off the end alternated between left and right for the three demonstrations. For the box with the button, the experimenter placed the box on the table so that the button was facing the infant. She then tried to push the button with the wand but missed all three times. For each attempt, she lifted the wand and slowly moved it toward the button but missed it by hitting slightly above, below, and to the right of the button. For the demonstration with the horizontal prong and loop object pair, the experimenter placed the prong device facing her, on her left hand side. This was done so that the infant could get a clear view of the demonstration. The experimenter picked up the loop and attempted to hang it on the prong, but "accidently" missed all three times. For the cup and beads trial, the experimenter placed the cup in front of her on the table and the beads just next to it. She then picked up the beads and attempted to place them inside the cup, but missed all three times. For the demonstration with the plastic square and wooden dowel, the experimenter first placed the objects on the table in front of her. She then picked up the plastic square, and using both hands, attempted to place the square onto the dowel, but missed all three times. After the demonstrations for each novel object pair, the experimenter offered the objects to the child and said, "Now it's your turn." Children's successful completion of the experimenter's intended action was coded.

Belief tasks

The True and False Belief tasks were adapted from Buttelmann and colleagues' (2009) original study. The main adaptation was that the tasks were completed at a table as opposed to on the floor, so that children would not have to change locations throughout the testing session. The two wooden boxes were placed on the table in front of the child, equidistant from one another and from the child. The task began with E1 saying to the child, "Wow, look at these boxes" as she opened and tilted each box so that the child could see that they were in fact empty. E1 then announced that she was going to find a toy for her and the child, and left the room. Then, in E1's absence, E2 demonstrated to the child how to lock and unlock the boxes using the wooden pins. Next, E2 then gave the child the opportunity to lock and unlock the boxes. The child had to lock and unlock each box twice before moving forward with the task. Once E2 returned to her spot next to the child, E1 re-entered the room with a plastic caterpillar. She sat across from the child and E2 and showed them the toy. E1 then offered to play with the child by rolling the caterpillar across the table and around the boxes. This play period lasted approximately 1 to 2 minutes, or until the child lost interest in the caterpillar. Next, E1 suggested that they put the caterpillar inside one of the boxes. She then put the toy into the box and closed it. E1 then exclaimed, "Oops! I forgot my keys outside. I'm going to go get them, I'll be right back" and then left the room. In E1's absence, E2 then said to the child, "Shhhh, let's play a trick on [E1's name]" as she proceeded to take the toy from box 1 and place it into box 2. Next she said, "Now let's lock the box" as she placed the pin in each box. E1 then returned to the testing room and stood in front of the two boxes on the table and said "So". She then tried to open the box where she had originally placed her toy. When she was unsuccessful at opening the box, she turned toward the child and said, "Hmmm?" with a confused facial expression. At this point, she looked down slightly, with her gaze in between the two boxes. If the child did not proceed to touch or unlock a box, E2 prompted the child to help E1. When the child unlocked or tried to unlock a box, E1 thanked the child for his or her effort. This task was administered to measure children's understanding of false beliefs. Therefore, if the child tried to open the box that now contained the toy (the full box), then it was assumed that he or she understood that E1 did not know that her toy had been secretly moved to box 2. That is, the child understood that E1 had a false belief about the toy's location. The True Belief task was administered similarly except that the experimenter remained present when the object was moved.

Desire task

The Desire task was adapted from Repacholi and Gopnik (1997) and included two conditions: a Match and a Mismatch condition. In both conditions, E1 placed two plastic containers on a tray, equidistant from one another and from the child. For the match condition, children were presented with crackers and broccoli. The location of the food was counterbalanced so that half of the time the crackers were on the child's left, and the other half of the time, on the child's

right. First, E1 invited the child to try the food. The child was allowed adequate time to taste one or both foods. During this time, the experimenter observed which food the child tasted first, and therefore was assumed to prefer. Then, E1 placed the containers in front of, but just outside of the child's reach, and tasted the food from each container. In the Match condition, the experimenter expressed pleasure when tasting the food the child preferred (usually crackers), and disgust when tasting the food that the child liked least (usually broccoli). That is, the experimenter said "Mmmm Crackers, mmm", and then "Eww Broccoli, eww", while displaying appropriate facial expressions. Next, the experimenter placed the containers in front of the child and said, "Can I have some?" as she placed her hand, palm up, in between the two containers. The experimenter always waited for the child to have nothing in their hands before making her request. The same procedure was used in the Mismatch condition, except that the experimenter demonstrated pleasure when tasting the food the child appeared not to prefer, and disgust when tasting the food that the child preferred. Children's choice of food to give to the experimenter was observed. If children gave the experimenter the food they preferred in the Mismatch condition, it was assumed that they had an understanding of conflicting desires.

2.4 | Coding and reliability

For the interactive tasks, participants' behaviors during each task were coded off-line using video recordings of the sessions. To measure inter-rater reliability, a second coder who was blind to the hypotheses coded the videos; Cohen's Kappa agreement values ranged from $k = .74$ to $k = .97$, showing a high degree of agreement. For the VOE tasks, infants' looking times at the scene during each task were coded offline using INTERACT 8.0 (Mangold, 2010). To establish inter-rater reliability, an independent observer coded a minimum of 25% of the data. Using Pearson product-moment correlations to compare overall looking time at the scene, the mean inter-observer reliability was calculated. In all cases, reliability was above $r = .90$ ($ps < .001$).

3 | RESULTS

Correlations computed between the pass/fail scores on both set of tasks and receptive vocabulary size ($M = 51.93$, $SD = 24.09$) revealed no link between these variables for any of the ToM constructs (all $rs < .16$, ns). To examine the main hypothesis, that is, whether or not

infants' performance on the interactive ToM tasks was related to infants' performance on the VOE ToM tasks, only infants who completed both tasks for each of the ToM abilities were included in the analyses. Therefore, the sample size for each set of analyses varies from 25 to 46. In addition, to ensure that infants were attentive during the VOE tasks, infants' looking time during each demonstration phase was examined to control for potential fatigue effects. The average percentage of time infants watched the demonstration phase of the Intention task was $M = 99\%$ ($SD = .03$, range = 76%–100%). Similarly, the average percentage of time infants watched the True Belief test demonstration phase was $M = 98\%$ ($SD = .05$, range = 78%–100%), and $M = 97\%$ ($SD = .05$, range = 73%–100%) for the False Belief task. Lastly, the percentage of looking time at the demonstration phase of the Desire task was $M = 99\%$ ($SD = .03$, range = 79%–100%). These results suggest high attention during the VOE tasks in the final sample.

Preliminary analyses revealed that the looking time measures for the False Belief, True Belief, Intention, and Desire tasks were not normally distributed, and therefore an additive (+1) log10 transformation was applied. Following these adjustments, the data were normally distributed, thereby meeting the normality assumption for parametric statistical tests. As the results from analyses on both raw and transformed scores revealed the same findings, only those computed with the original raw scores are reported. Table 1 shows the average looking time at the incongruent and congruent scenes, as well as the average proportion score on the VOE tasks for the entire sample. Infants looked significantly longer at the incongruent trial on the Intention and Desire tasks. They also tended to look longer at the incongruent scene in the case of True Belief, but looked equally long at the incongruent and congruent test trials in the False Belief task. A preference score was then calculated, based on infants' looking time during the incongruent test trial divided by infants' total looking time during the incongruent and congruent test trials combined. A score above 50% reflected longer looking time at the incongruent scene. As seen in Table 1, none of these proportion scores were above chance (50%). As in the original study (Buttelmann et al., 2009), a 2 (FB vs. TB task) × 2 (empty box vs. full box) repeated-measure analysis of variance contrasting looking times on the False Belief and True Belief tasks was conducted. Results revealed a Condition × Trial interaction ($F(1, 34) = 7.73$, $p < .009$) with longer looking time at the empty box than at the full box in the True Belief condition, but no difference in the False Belief condition. In order to analyze individual differences in performance, participants were classified as having passed or failed each task. To establish a passing criterion, participants who looked longer

TABLE 1 Mean looking times and proportion scores on the incongruent and congruent trials in VOE tasks

Task	Incongruent scene		Congruent scene					Proportion score	
	M	SD	M	SD	t (df)	p	d	M	SD
Intention	7.13	4.72	5.32	2.95	−2.34 (32)	.02	−.46	.54	.18
True Belief	8.66	6.25	6.29	4.26	−1.97 (35)	.06	−.44	.56	.22
Desire	8.99	1.14	8.41	2.03	−2.11 (45)	.04	−.35	.52	.07
False Belief	8.53	6.35	7.45	5.74	.96 (42)	.34	−.18	.53	.20

TABLE 2 Proportion of looking time in the incongruent trials in the VOE tasks as a function of performance on the interactive tasks

Interactive task	Pass M (SD)	Fail M (SD)	T	p	d
VOE task					
Intention	0.56 (0.19)	0.52 (0.16)	−0.68	0.50	−0.23
True Belief	0.61 (0.24)	0.52 (0.24)	−0.91	0.37	−0.38
Desire	0.55 (0.08)	0.51 (0.06)	−1.56	0.13	−0.57
False Belief	0.54 (0.22)	0.58 (0.19)	0.44	0.66	0.19

Note. Ratio was calculated as time spent looking at the incongruent trials as a proportion over the total looking time for that task.

than 50% at the incongruent trial were classified as passing the task. The results showed that 55.9% of infants passed the intention task, 50% passed the true belief task, 47.8% passed the desire task, and 46.5% passed the false belief task. The proportion of children who passed these tasks was at chance level (50%; binomial tests: ps = .36 to 1.00).

With regard to the interactive tasks, analyses revealed that 42.9% of the infants passed the intention task (pass criterion set at 75% correct trials), 48.3% passed the true belief task, 43.5% passed the desire task, and 36.6% passed the false belief task. None of these proportions were above chance (50%) as assessed by binomial tests (ps = .44 to 1.00). As in the original false belief study, we compared performance across the True and False Belief conditions and the box children chose did not differ significantly between conditions, $\chi2$ (1, N = 25) = .43, p = .51). Finally, some children required some prompting from the experimenter but only three children refused to choose a box in the FB task and five in the TB task.

To test our main hypothesis concerning the convergence across tasks, independent samples t tests were computed to compare the performance on each VOE task (proportion score) between those who passed versus those who failed the corresponding interactive task. As shown in Table 2, results comparing performance across tasks revealed that infants' performance on the interactive task was unrelated to their preference score in the equivalent VOE task, for each pair of tasks. In other words, infants who were successful on a given interactive task did not look longer during the incongruent trial than those who failed. Because the intention task was measured on a continuous scale, we also computed the proportion of correct actions to the VOE proportion looking time with similar results as with the dichotomous variable (r = −.07, p = .68).

4 | DISCUSSION

In this study, 18-month-olds' understanding of intention, desire, true, and false belief was assessed with two batteries of tasks, one based on the VOE paradigm and the other on the active, prompted-action method. These tasks are considered implicit as no verbal skills are required to succeed, as confirmed by the lack of relation with receptive vocabulary, which replicates previous research (Olineck & Poulin-Dubois, 2009; Poulin-Dubois & Yott, 2014; Yott & Poulin-Dubois, 2016). By using a within-subject design to assess infants' implicit ToM

skills, we aimed to address the nature (shallow or deep) of infants' processing of events that involve agents acting on objects. We reasoned that if these tasks are, as argued by the proponents of the mentalistic view, measuring infants' representations of mental states, then one would expect that performance on the VOE tasks would be related to performance on the interactive tasks (Scott et al., 2015). In contrast, if performance on the VOE tasks is mainly driven by more rudimentary abilities, and by "prosocial" skills in interactive tasks, then there is no reason to expect consistency across procedures.

Overall, infants performed as previously reported on the VOE tasks, looking longer at the incongruent trial than at the congruent trial, except in the case of the false belief task. The pattern of results in the case of the true belief task was in the expected direction but did not reach statistical significance. Given that the order of administration of the VOE belief tasks was counterbalanced, fatigue effects cannot explain infants' poor performance, particularly on the false belief task. We believe that it is most likely due to the within-subject design that was required to assess inter-task coherence (congruent and incongruent test trials) as participants most likely recognized the display and the experimenter from the previous testing session. The fact that the participants were older than in the original experiment by Onishi and Baillargeon (18 vs. 15 months) could also have affected performance but it is unlikely as a similar null result was recently reported in a study combining the present sample with a sample of 14-month-olds (Yott & Poulin-Dubois, 2016). Importantly, another failed replication has recently been reported in 18-month-olds using a similar within-subject design, whereas replications have been reported when a between-subject design is used (Powell, Hobbs, Bardis, & Carey, 2017; Yott & Poulin-Dubois, 2012). Finally, it is worth noting that two groups of researchers have recently reported low passing rates on anticipatory looking false belief tasks in infants, toddlers, and adults (Dörrenberg, Rakoczy, & Liszkowski, 2017; Kulke & Rakoczy, 2017; Sodian et al., 2016; Thoermer et al., 2012). Limited statistical power could also account for the current findings, given the constraint that children needed to complete a pair of tasks in order to be included in each set of inter-task analyses. However, the current sample sizes are equivalent to those on which many rich interpretations of infants' behaviors are based in the extant literature on false belief. Clearly, performance on the VOE tasks, particularly the false belief VOE task, appears to be unstable and directly affected by context-specific variables.

With regard to the other battery of tasks based on an active, prompted-action paradigm, performance also varied across tasks, with

replication only observed in the case of the behavioral enactment task used to measure intention understanding. In contrast to the original study (Repacholi & Gopnik 2007), and in line with many studies, infants were at chance on the desire task where they had to offer the experimenter food that they themselves disliked (Carlson et al., 2004; Chiarella et al., 2013; Hobbs & Spelke, 2015; Sodian et al., 2016; Wright & Poulin-Dubois, 2012). Thus, the success rate on the original desire task (69%) has been difficult to replicate in a large number of recent studies, even in older infants. Finally, in one of the first attempts to replicate the false belief task based on helping the experimenter find a toy by taking into account her beliefs (Buttelmann et al., 2009), our participants performed very poorly. Given the very high attrition rate reported in the original study and the need for parents to often prompt the child to walk toward the display to help the experimenter, we administered the task at a table. Despite the significant rise in infants' compliance, this methodological change decreased the amount of time available to make a reasoned decision due to the accessibility of the boxes. It is worth mentioning that two recent attempts to replicate the original findings have been successful when strictly following the original protocol, with the success rate varying from 62% to 82% (Powell et al., 2017; Priewasser, Rafetseder, Gargitter, & Perner, 2017).

Despite the fact that performance tended to be poorer than that observed with the original, between-subject design, none of the tasks yielded a floor effect and there was sufficient individual variability to allow a comparison of performance across each pair of tasks measuring the same ToM construct. The findings were straightforward: there were no significant relations observed between performance on the two sets of tasks for all four ToM constructs that were measured. What do these findings suggest about the depth of infants' theory of mind? At first glance, this pattern of performance seems to challenge the mentalistic view of infants' performance on implicit ToM tasks. According to the original proposal, the capacity to attribute mental states, such as false beliefs, to agents emerges in the first year of life and that the high task demands of standard theory of mind tasks mask this infants' psychological reasoning system (Baillargeon et al., 2010). In a revised mentalist account, it has been proposed that infants will not always pass non-elicited response tasks because it depends on their ability to cope with the demands imposed by the situation. For example, in the VOE task, no response selection or response inhibition are required to succeed. In contrast, in the prompted-action or helping task (a hybrid between spontaneous- and elicited-response tasks) it is argued that a response selection process is activated but no response inhibition (Baillargeon et al., 2015). Thus, a proponent of the mentalistic view might argue that the infants who passed the VOE task might have failed the helping task not because they do not possess false belief understanding but because they were unable to execute the appropriate response (open the box that contains the toy). In other words, even if strict replications had been attempted, lack of inter-task convergence would still be expected. In the present study, methodological details might explain the failure to replicate the original tasks as well as the failure to observe convergent validity across tasks. The other possibility is that the poor replications and lack of inter-task correlations put into question the robustness of theory of mind

competence in infancy. The present findings do not allow teasing apart these two possibilities. No doubt, the present findings would benefit from a replication with a different set of theory of mind tasks. For example, infants' performance on an anticipatory looking task could be compared with performance on another interactive task based on helping (Southgate et al., 2007; Southgate et al., 2010). Interestingly, such a comparison has recently been reported for false belief understanding and a lack of inter-task coherence between the anticipatory looking and helping tasks was observed in 24-month-old toddlers (Dörrenberg et al., 2017). Furthermore, even scores on spontaneous-responses tasks testing implicit false belief with equivalent processing demands (e.g., anticipatory looking) have also yielded no convergent validity in a very large sample of adults (Kulke & Rakoczy, 2017).

We predicted that if infants already possess a mature ToM, then performance on spontaneous-response or prompted-action tasks should show the same level of stable individual differences as what has been reported in older children with standard, elicited-response tasks. For example, strong inter-task convergence has been reported for false belief understanding in preschool children (Carlson & Moses, 2001; Thoermer et al., 2012). Although the observed lack of convergent validity appears to pose a challenge for the mentalistic account, it does not inform about which of the specific lean accounts best explains infants' behaviors in implicit ToM tasks. Some lean views have proposed that infants respond to the perceptual novelty of the test events or to new agent-object relations in the VOE tasks or to violations of learned behavioral rules (Heyes, 2014; Ruffman, 2014). These skills are assumed to provide the building blocks for the development of an implicit understanding of behavior that is expressed in success on ToM tasks. Only later, with the development of language skills and executive functions, do children develop an explicit understanding of mental states. Lean interpretations of the active or imitation paradigms have also rejected a mentalistic interpretation by arguing that these procedures are assessing an understanding of social situations (e.g., playing tricks in false belief task) or the object-directed behaviors of the agent and not mindreading skills (Allen, 2015; Priewasser et al., 2017).

We believe that null results can be critical in science when they counter a predicted effect. However, given that the interpretation of null results always poses a challenge, a number of alternative, leaner interpretations need to be considered for the present findings. One interpretation is that examining how well performance on one task relates to performance on another task at the same age (stability in the relative rank within a group) is risky as there is so much "noise" in tests measuring infant cognition, in comparison to those used with older children and adults. However, convergent validity has been demonstrated in a wide range of infant abilities, including motor skills, language, and attention (Campbell et al., 2013; Colombo & Mitchell, 2014; Pan, Rowe, Spier, & Tamis-LeMonda, 2004). In conclusion, the current study used a within-subject design in an effort to clarify the underlying basis for infants' performance in tasks that have been designed to assess ToM abilities in preverbal children. By testing infants on both passive looking and active action paradigms we aimed to determine if, despite the different task demands of these procedures,

mindreading skills are involved in infants' ability to predict agents' behaviors across all these scenarios. The current findings show that there is a lack of consistency across these measures, providing a new way to interpret infants' performance on tasks based on implicit spontaneous responses. Taken together, our results suggest that the attribution of abstract mindreading skills to infants might be premature and that additional research based on innovative approaches like the one reported in the present study are needed to clarify the current debate on the depth of infants' theory of mind.

ACKNOWLEDGEMENTS

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.

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

Aschersleben, G., Hofer, T., & Jovanovic, B. (2008). The link between infant attention to goal-directed action and later theory of mind abilities. *Developmental Science*, *11*, 862–868.

Baillargeon, R., Scott, R.M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, *67*, 159–186.

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

Baillargeon, R., Scott, R.M., He, Z., Sloane, S., Setoh, P., Jin, K., ... Bian, L. (2015) Psychological and sociomoral reasoning in infancy. In M. Mikulincer & P.R. Shaver (Eds.), E. Borgida & J.A. Bargh (Assoc. Eds.), *APA handbook of personality and social psychology: Vol. 1. Attitudes and social cognition* (pp. 79–150). Washington, DC: American Psychological Association.

Bellagamba, F., Camaioni, L., & Colonnesi, C. (2006). Change in children's understanding of others' intentional actions. *Developmental Science*, *9*, 182–188.

Brink, K.A., Lane, J.D., & Wellman, H.M. (2015). Developmental pathways for social understanding: Linking social cognition to social contexts. *Frontiers in Psychology*, *6*, 719.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*, 337–342.

Butterfill, S.A., & Apperly, I.A. (2013). How to construct a minimal theory of mind. *Mind & Language*, *28*, 606–637.

Campbell, S.K., Zawacki, L., Rankin, K.M., Yoder, J.C., Shapiro, N., Li, Z., & White-Traut, R. (2013). Concurrent validity of the TIMP and the Bayley III scales at 6 weeks corrected age. *Pediatric Physical Therapy*, *25*, 395–401.

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

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

Chiarella, S.S., Kristen, S., Poulin-Dubois, D., & Sodian, B. (2013). Concurrent relations between perspective-taking skills, desire understanding, and internal-state vocabulary. *Journal of Cognition and Development*, *14*, 480–498.

Clements, W.A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377–395.

Colombo, J., & Mitchell, D.W. (2014). Individual differences in early visual attention: Fixation time and information processing. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, and prediction* (pp. 193–229). New York: Psychology Press.

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

Fenson, L., Pethick, S., Renda, C., Cox, J.L., Dale, P.S., & Reznick, J.S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, *21*, 95–115.

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

He, Z., Bolz, M., & Baillargeon, R. (2011). False-belief understanding in 2.5-year-olds: Evidence from violation of expectation change of location and unexpected-contents tasks. *Developmental Science*, *14*, 292–305.

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

Hobbs, K., & Spelke, E.S. (2015). Goal attributions and instrumental helping at 14 and 24 months of age. *Cognition*, *142*, 44–59.

Hughes, C., & Ensor, R. (2005). Executive function and theory of mind in 2 year olds: A family affair? *Developmental Neuropsychology*, *28*, 645–668.

Knudsen, B., & Liszkowski, U. (2012a). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy*, *17*, 672–691.

Knudsen, B., & Liszkowski, U. (2012b). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science*, *15*, 113–122.

Kovács, Á.M., Téglás, E., & Endress, A.D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*, 1830–1834.

Kulke, L., & Rakoczy, H. (2017, April). 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.

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.

Mangold, P. (2010). *Interact user guide, V. 8.0*. Arnstorf, Germany: Mangold International.

Meltzoff, A.N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*, 838–850.

Olineck, K.M., & Poulin-Dubois, D. (2009). Infants' understanding of intention from 10 to 14 months: Interrelations among violation of expectancy and imitation tasks. *Infant Behavior and Development*, *32*, 404–415.

Onishi, K.H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–258.

Pan, B.A., Rowe, M.L., Spier, E., & Tamis-LeMonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: Concurrent and predictive validity of three sources of data. *Journal of Child Language*, *31*, 587–608.

Perner, J. (2010). Who took the cog out of cognitive science? Mentalism in an era of anti-cognitivism. In P.A. Frensch (Ed.), *Cognition and neuropsychology: International perspectives on psychological science* (pp. 241–261). New York: Psychology Press.

Phillips, A.T., & Wellman, H.M. (2005). Infants' understanding of object-directed action. *Cognition*, 98, 137–155.

Phillips, J., Ong, D.C., Surtees, A.D., Xin, Y., Williams, S., Saxe, R., & Frank, M.C. (2015). A second look at automatic theory of mind: Reconsidering Kovács, Téglás, and Endress (2010). *Psychological Science*, 26, 1353–1367.

Poulin-Dubois, D. (1999). Infants' distinction between animate and inanimate objects: The origins of naive psychology. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life* (pp. 257–280). Mahwah, NJ: Lawrence Erlbaum Associates.

Poulin-Dubois, D., Brooker, I., & Chow, V. (2009). The developmental origins of naïve psychology in infancy. *Advances in Child Development and Behavior*, 37, 55–104.

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, 87–99.

Poulin-Dubois, D., & Yott, J. (2014). Executive functions and theory of mind understanding in young children: A reciprocal relation? *Psychologie Française*, 59, 59–69.

Powell, L., Hobbs, K., Bardis, A., & Carey, S. (2017, April). Investigating the reproducibility of implicit theory of mind tasks with different levels of representational demands. In H. Rokoczy & 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: Belief or interest? Manuscript submitted for publication.

Rakoczy, H., Bergfeld, D., Schwarz, I., & Fizke, E. (2014). Explicit theory of mind is even more unified than previously assumed: Belief ascription and understanding aspectuality emerge together in development. *Child Development*, 86, 486–502.

Repacholi, B.M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, 33, 12–21.

Rice, K., & Redcay, E. (2015). Spontaneous mentalizing captures variability in the cortical thickness of social brain regions. *Social Cognitive and Affective Neuroscience*, 10, 327–334.

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review*, 34, 265–293.

Ruffman, T., & Perner, J. (2005). Do infants really understand false belief? *Trends in Cognitive Sciences*, 9, 462–463.

Scott, R.M. (2017). The developmental origins of false-belief understanding. *Current Directions in Psychological Science*, 26, 68–74.

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

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

Scott, R.M., Richman, J.C., & Baillargeon, R. (2015). Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology*, 82, 32–56.

Scott, R.M., & Roby, E. (2015). Processing demands impact 3-year-olds' performance in a spontaneous-response task: New evidence for the processing-load account of early false-belief understanding. *PLoS ONE*, 10, e0142405.

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

Sodian, B. (2011). Theory of mind in infancy. *Child Development Perspectives*, 5, 39–43.

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.

Song, H.J., Onishi, K.H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition*, 109, 295–315.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, 13, 907–912.

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

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18, 580–586.

Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology*, 30, 30–44.

Surtees, A.D., Butterfill, S.A., & Apperly, I.A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology*, 30, 75–86.

Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology*, 30, 172–187.

Trudeau, N., Frank, I., & Poulin-Dubois, D. (1999). Une adaptation en français québécois du MacArthur Communicative Development Inventory [A Quebec-French adaptation of the MacArthur Communicative Development Inventory]. *La revue d'orthophonie et d'audiologie*, 23, 61–73.

Wellman, H.M. (2014). *Making minds: How theory of mind develops*. Oxford: Oxford University Press.

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

Wellman, H.M., Lopez-Duran, S., LaBounty, J., & Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Developmental Psychology*, 44, 618–623.

Wellman, H.M., Phillips, A.T., Dunphy-Lelii, S., & LaLonde, N. (2004). Infant social attention predicts preschool social cognition. *Developmental Science*, 7, 283–288.

Woodward, A.L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69 (1), 1–34.

Wright, K., & Poulin-Dubois, D. (2012). Modified checklist for autism in toddlers (M-CHAT) screening at 18 months of age predicts concurrent understanding of desires, word learning and expressive vocabulary. *Research in Autism Spectrum Disorders*, 6, 184–192.

Yamaguchi, M., Kuhlmeier, V.A., Wynn, K., & van Marle, K. (2009). Continuity in social cognition from infancy to childhood. *Developmental Science*, 12, 746–752.

Yott, J., & Poulin-Dubois, D. (2012). Breaking the rules: Do infants have a true understanding of false belief? *British Journal of Developmental Psychology*, 30, 156–171.

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.