How Sophisticated Is Infants' Theory of Mind?

Rose M. Scott

University of California, Merced

Erin Roby

New York University School of Medicine

Renée Baillargeon

University of Illinois at Urbana-Champaign

Citation:

Scott, R. M., Roby, E., & Baillargeon, R. (in press). How sophisticated is infants' theory of mind? To appear in O. Houdé & G. Borst (Eds.-in-chief), *Cambridge handbook of cognitive development*. Cambridge, England: Cambridge University Press.

I. Introduction

Imagine the following scenario (inspired by true events): We are grocery shopping with our friend Anita, who notices a woman down the aisle and waves and smiles at her. The woman looks confused and walks away. Perplexed, Anita tells us that she spent the previous day training the woman, a new employee at their firm, and is surprised by her behavior because she thought they had established a friendly rapport. We suggest a possible explanation: Perhaps the woman did not recognize Anita outside of work. Later at the checkout line, we notice the woman standing with someone who is obviously her identical twin. In light of this new information, we revise our initial explanation: Anita must have waved at her trainee's twin, falsely believing her to be her trainee, and the twin responded with confusion because she had never met Anita before. Anita then joins us at the checkout, one of the twins waves happily to her, and the mystery is resolved with smiles all around.

As this everyday example illustrates, adults routinely attempt to infer others' mental states to make sense of their actions and interactions. Variously known as psychological reasoning, mental-state reasoning, mentalizing, mindreading, or theory of mind, this ability plays a vital role in everyday social life and is associated with a wide range of positive developmental outcomes, from improved cooperation to better academic performance (Butler, 2013; Cowell et al., 2017; Imuta, Henry, Slaughter, Selcuk, & Ruffman, 2016; Takagishi, Kameshima, Schug, Koizumi, & Yamaghisi, 2010). Considerable research has thus focused on when and how this important ability develops.

Over the past 25 years, our understanding of early psychology reasoning has undergone several substantial shifts, each prompted by new findings that revealed previously unsuspected competencies (for reviews, see Baillargeon et al., 2015; Baillargeon, Scott, & Bian, 2016; Scott, 2017; Scott & Baillargeon, 2017). In this chapter, we outline the successive accounts of early psychological reasoning corresponding to these shifts. Section II describes earlier accounts, including constructivist, teleological, and conceptual-change accounts. Section III introduces the current debate between two-system and one-system accounts. Section IV first reviews challenges to "signature limits" of early psychological reasoning proposed by two-system accounts, and then considers in detail one limit that has until now received scant attention, cognitive encapsulation. Finally, Section V outlines future research directions.

II. Earlier Accounts

A. Constructivist and Teleological Accounts

In the early 1990s, many researchers believed, following the Piagetian tradition, that infants in the first year of life had no capacity for psychological reasoning and lacked any understanding of others as agents (Butterworth & Jarrett, 1991; Tomasello, 1999). However, new findings suggested that young infants could make sense of familiar actions by human agents (e.g., reaching for and grasping an object), though not by non-human agents (Kamewari, Kato, Kanda, Ishiguro, & Hiraki, 2005; Meltzoff, 1995; Woodward, 1998, 1999). These findings led to *constructivist* or experience-driven accounts of early psychological reasoning (Woodward, 2005; Woodward, Sommerville, & Guajardo, 2001). According to these accounts, as infants learned to act on objects, they became able to understand similar actions by similar agents. Over time, through the application of comparison and abstraction processes to the representations of their own actions and those of other agents, infants gradually constructed a more abstract understanding of intentional action, which could then be applied, broadly and flexibly, to novel actions and novel agents.

In time, constructivist accounts were called into question by new evidence that young infants could make sense of intentional actions by novel non-human agents, provided they received sufficient evidence to identify them as agents—specifically, evidence that these novel entities had internal control over their actions (Csibra, 2008; Csibra, Gergely, Bíró, Koós, & Brockbank, 1999; Johnson, Shimizu, & Ok, 2007; Luo & Baillargeon, 2005). These findings supported *teleological* accounts, which suggested that infants possessed a primitive cognitive system that enabled them to reason about the actions of any agents, whether human or non-human; this system was thought to be non-mentalistic in nature and, as such, to be qualitatively different from the mentalistic psychological-reasoning system of older children and adults (Csibra, Bíró, Koós, & Gergely, 2003; Gergely & Csibra, 2003). According to such accounts, infants reasoned not about agents' mental states per se, but rather about physical variables related to these mental states. When watching an agent act in a scene, infants generated a teleological explanation that specified the physical layout of the scene, the agent's actions, and the end-state the agent achieved. This teleological explanation, together with a core principle of *rationality* (all other things being equal, agents act rationally; Dennett, 1987; Gergely, Nádasdy, Csibra, & Bíró, 1995), was sufficient to allow infants to predict what the agent would do when the layout changed. For example, if agent-1 repeatedly jumped over an obstacle and then approached agent-2 on the far side of the obstacle, infants would expect agent-1 to travel to agent-2 in a straight line once the obstacle was removed (i.e., rational agents expend as little effort as possible to achieve end-states; Csibra et al., 1999; Gergely et al., 1995). Proponents of the teleological view argued that with development, physical variables gradually became incorporated into a mentalistic system that made sense of agents' actions in terms of intentions and other mental states (Gergely & Csibra, 2003).

Implicit in teleological accounts was the assumption that infants were fundamentally egocentric: Because they attended only to physical variables, they could not distinguish between their own representation of a layout and that of an agent. Thus, if infants could see two objects but an agent could see only one of them, infants would be unable to distinguish between their representation of the scene (two objects) and that of the agent (one object). In time, however, new evidence indicated that infants did recognize that an agent could be ignorant about information that was available to them, but not to the agent. For example, they understood that an agent would be ignorant about an object that was hidden from her by a barrier (Kim & Song, 2015; Luo & Baillargeon, 2007) or about an event she had not witnessed (Liszkowski, Carpenter, & Tomasello, 2008; Tomasello & Haberl, 2003). In such cases, infants not only kept track of what the agent could see or had seen, but also non-egocentrically adopted the agent's representation of the scene to predict and interpret her actions.

Together, these findings suggested that infants might have something akin to a mentalistic system for reasoning about agents' actions. Further evidence for this conclusion came from investigations focusing on another facet of early social cognition, sociomoral reasoning. These investigations provided new evidence that infants (a) could identify novel non-human agents (Hamlin, Wynn, & Bloom, 2007; Margoni, Baillargeon & Surian, 2018), (b) could keep track, non-egocentrically, of what knowledge agents possessed or lacked about a scene (Meristo & Surian, 2013; Sloane, Baillargeon, & Premack, 2012), (c) could attribute to agents a wide range of positive (e.g., helping, comforting, giving) and negative (hindering, hitting, stealing) intentions (Hamlin et al., 2007; Hamlin & Wynn, 2011; Jin, Houston, Baillargeon, Groh, & Roisman, 2018; Margoni et al., 2018), and (d) could distinguish between positive and negative intentions even if agents failed in their attempts to carry out these intentions (Dunfield & Kuhlmeier, 2010; Hamlin, 2013).

B. Conceptual-Change Accounts

Despite the wealth of evidence that infants could reason about agents' mental states, it remained unclear how similar infants' mentalistic system was to that of older children and adults. Infants were clearly able to understand simple mental states such as motivational states (e.g., intentions) and epistemic states (e.g., ignorance), but could they also understand more complex mental states such as false beliefs? According to *conceptual-change* accounts, the answer to this question had to be no, because there was ample evidence suggesting that children younger than 4 years of age could not represent false beliefs. This evidence came primarily from traditional or elicited-prediction tasks, which require children to answer a direct question about the likely behavior of a mistaken agent (Baron-Cohen, Leslie, & Frith, 1985; Gopnik & Astington, 1988; Perner, Leekam, & Wimmer, 1987; Wimmer & Perner, 1983). In one well-known transfer-of-location task (Baron-Cohen et al., 1985), children see Sally place a marble in a box; in her absence, the marble is moved to a nearby basket. Children are then asked where Sally will look for the marble when she returns. Beginning at around age 4, children correctly indicate that Sally will look in the box, where she falsely believes the marble is located. In contrast, younger children incorrectly indicate that Sally will look in the basket, the marble's true location, suggesting an inability to appreciate Sally's false belief. This developmental pattern was widely replicated (Liu, Wellman, Tardif, & Sabbagh, 2008; Wellman, Cross, & Watson, 2001), leading conceptual-change theorists to conclude that a fundamental leap took place in psychological reasoning at about age 4, when children began to grasp the representational nature of the mind. False-belief understanding thus came to be viewed as a major developmental milestone that was not achieved until the preschool years and heralded a more sophisticated form of psychological reasoning (Carlson & Moses, 2001; de Villiers & de Villiers, 2003; Gopnik & Wellman, 1994; Perner, 1991).

Conceptual-change accounts were eventually challenged, however, by new evidence suggesting that children under age 4 can demonstrate false-belief understanding when tested with tasks other than traditional tasks. Non-traditional tasks do not require answering a direct question about the likely behavior of a mistaken agent (e.g., “Where will Sally look for her marble?”), but instead use alternative ways of assessing children’s understanding of the agent’s false belief. The first non-traditional task administered to infants (Onishi & Baillargeon, 2005) used the violation-of-expectation method, which takes advantage of infants’ natural tendency to look longer at events that violate, as opposed to confirm, their expectations. In the task, 15-month-olds first saw an agent hide a toy in box-A as opposed to box-B. Next, infants received one of several belief-induction trials in which the agent came to hold either a true or false belief about the toy’s location. In the subsequent test trial, the agent reached into either box-A or box-B and then paused. Infants expected the agent to reach into whichever box she believed contained the toy, regardless of whether her belief was true or false, and they detected a violation if she reached into the other box instead.

Since the publication of these results, over 30 reports have produced positive evidence of early false-belied understanding in 6- to 36-month-olds (for a partial review, see Scott & Baillargeon, 2017).[1] These reports have used 11 different non-traditional methods, including (a) behavioral spontaneous-response tasks (violation-of-expectation, anticipatory-looking, preferential-looking, anticipatory-pointing, and affective-response tasks; Knudsen & Liszkowski, 2012; Moll, Khalulyen, & Moffett, 2017; Onishi & Baillargeon, 2005; Scott, He, Baillargeon, & Cummins, 2012; Southgate, Senju, & Csibra, 2007), (b) behavioral interactive or elicited-intervention tasks (helping and referential-communication tasks; Buttelmann, Carpenter, & Tomasello, 2009; Southgate, Chevallier, & Csibra, 2010), and (c) neural spontaneous-response tasks (action-prediction, sustained-representation, belief-processing, and semantic-incongruity tasks; Forgács et al., 2019; Hyde, Simon, Ting, & Nikolaeva, 2018; Kampis, Parise, Csibra, &

Kovács, 2015; Southgate & Vernetti, 2014). Non-traditional methods have yielded positive findings not only in Western cultures, as cited above, but also in non-Western cultures (Barrett et al., 2013).

III. More Recent Accounts

The converging findings from non-traditional tasks suggest that some capacity for reasoning about false beliefs is already present early in life. However, there is currently a heated debate about the exact nature of infants' false-belief understanding and psychological reasoning more generally. This debate primarily involves *two-system* accounts, which assume that traditional and non-traditional tasks are carried out by different psychological-reasoning systems (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; Low, Apperly, Butterfill, & Rakoczy, 2016; Low & Watts, 2013), and *one-system* accounts, which assume that both types of tasks are carried out by the same psychological-reasoning system (Baillargeon et al., 2016; Carruthers, 2013, 2016; Scott & Baillargeon, 2017). Deflationary accounts, which assume that infants' psychological reasoning can be explained in non-mentalistic terms using low-level stimulus factors, associations, or behavioral rules (Heyes, 2014; Perner, 2010; Ruffman, 2014), are not reviewed here. As Carruthers (2018) recently stated, "the consensus among most researchers in the field is that there are now too many infant studies, using too many variations in materials and methods, for [such accounts] to be plausible" (p. 11351).

A. Two-System Accounts

According to *two-system* accounts, two distinct cognitive systems underlie human psychological reasoning (Apperly & Butterfill, 2009; Butterfill, & Apperly, 2013; Low et al., 2016; Low & Watts, 2013). The late-developing system, which emerges around 4 years of age, is non-automatic, slow, and flexible; it is capable of representing a wide range of false beliefs and enables correct responses in traditional tasks. The early-developing system, which emerges in infancy, is automatic, fast, and inflexible; it tracks belief-like states or “registrations” that are sufficient for success at most non-traditional tasks. For example, if an agent hides an object in one location and, in her absence, the object is moved to another location, the early-developing system can predict, by considering where the agent last encountered and registered the object, that upon her return she will search for it in its original (as opposed to its current) location. Thus, by tracking what information is available to agents about objects’ locations and properties, even if it becomes outdated as events unfold, the early-developing system makes possible a minimal form of false-belief understanding.

Due to its primitive nature, the early-developing system has several “signature limits”, beyond its failure in traditional tasks. One such limit is that although infants can represent false beliefs about the locations and properties of objects, as we just saw, they cannot represent false beliefs about objects’ identities: Because registrations are minimal representations, they do not “allow for a distinction between what is represented and how it is represented” (Apperly & Butterfill, 2009, p. 963). Thus, if an infant and an agent view the same object but hold different beliefs about what it *is* (e.g., the infant knows it is a tree branch, but the agent mistakenly believes it is a snake), the early-developing system cannot correctly predict the agent’s actions. Another signature limit is that because the early-developing system is “largely automatic and independent of central cognitive resources” (Low et al., 2016, p. 185), it has limited information-processing resources and cannot handle demanding situations in which predicting an agent’s actions requires reasoning about multiple interlocking mental states that interact causally (Low, Drummond, Walmsley, & Wang, 2014). According to two-system theorists, such a complex causal structure “places demands on working memory, attention, and executive function that are incompatible with automaticity" (Butterfill & Apperly, 2013, p. 629). Yet another signature limit, related to the last, is that the early-developing system is largely encapsulated from the rest of cognition. This makes it fast and efficient, but also inflexible and sharply limited in its ability to integrate information from other cognitive processes when reasoning about agents' actions (Butterfill & Apperly, 2013).

B. One-System Accounts

According to *one-system* accounts, a single system underlies human psychological reasoning (Baillargeon et al., 2016; Carruthers, 2013, 2016; Scott & Baillargeon, 2017; for earlier nativist accounts, see Leslie, 1987, 1994). This system emerges early in infancy, is mentalistic in nature, and is constrained by a principle of rationality, with corollaries of *consistency* (agents act in a manner consistent with their mental states) and *efficiency* (agents expend as little effort as possible when pursuing their goals) (Baillargeon et al., 2016). This principle is both descriptive and normative in nature: Infants use the principle, together with their representations of agents' mental states, not only to predict but also to *evaluate* agents' actions. Thus, infants are less likely to trust, learn from, or direct inquiries to agents who behave irrationally (Begus & Southgate, 2012; Koenig & Woodward, 2010; Poulin-Dubois, Brooker, & Polonia, 2011; Zmyj, Buttelmann, Carpenter, & Daum, 2010).

Because one-system theorists assume that infants' psychological reasoning is qualitatively similar to that of older children and adults, they dispute all of the signature limits proposed by two-system accounts. Furthermore, they assume that young children fail at traditional tasks mainly because these tasks pose greater processing difficulties. Various descriptions of these difficulties have been offered, focusing on limitations in young children's executive functioning (Leslie, Friedman, & German, 2004; Wang & Leslie, 2016), pragmatic understanding (Hansen, 2010; Rubio-Fernandez & Geurts, 2016), or both (Helming, Strickland, & Jacob, 2016; Westra &

Carruthers, 2017). Building on these efforts, recent descriptions of early difficulties in traditional tasks have tended to include a wider range of factors (Antilici, 2019; Setoh, Scott, & Baillargeon, 2016).

According to the *expanded-processing-demands* view, for example, success at the Sally-Anne task depends on several different processes (Setoh, Scott, et al., 2016). As children hear the story, they must comprehend it, and build and maintain a representation of Sally's false belief. When asked the test question ("Where will Sally look for her marble?"), they must interpret it as intended and decide how to respond. Because children typically begin this response-generation process by considering where the marble currently is, they must then inhibit that incorrect prepotent response to tap their representation of Sally's false belief. Finally, children must generate the correct response by naming or pointing to the marble's original location (e.g., the basket). According to the expanded-processing-demands view, young children may fail at the Sally-Anne task either because they lack sufficient skill at one or more of the processes involved (e.g., they lack sufficient inhibitory control to inhibit the incorrect prepotent response triggered by the test question) or because they lack sufficient information-processing resources to handle the total processing demands of the task. This view is consistent with findings that improvements in various cognitive capacities—including inhibitory control, working memory, and verbal ability—contribute to success at traditional tasks (Carlson & Moses, 2001; Devine & Hughes, 2014; Duh et al., 2016; Milligan, Astington, & Dack, 2007), as does practice at conversing about mental states (Hofmann et al., 2016; McAlister & Peterson, 2006; Ruffman, Slade, & Crowe, 2002; Wang & Su, 2009). Both cognitive advances and exposure to mental-state talk can help reduce the processing demands associated with answering direct questions about agents' false beliefs, thus enabling success at traditional tasks (see also Antilici, 2019).

IV. Evidence against Two-System Accounts

A. Recent Challenges

Do two-system or one-system accounts better describe early psychological reasoning? Over the past 10 years, evidence has begun to accumulate that casts doubt on two-system accounts. Below, we review four such sets of findings.

Two systems. Recent neuroscientific evidence provides no support for the claim that two systems with distinct neurological substrates and computational capacities underlie success in traditional and non-traditional tasks. First, investigations with adults, using either functional magnetic-resonance imaging (fMRI; Bardi, Desmet, Nijhof, Wiersema, & Brass, 2017) or functional near-infrared spectroscopy (fNIRS; Hyde, Aparicio Betancourt, & Simon, 2015), indicate that traditional and non-traditional tasks engage anatomically similar regions within the temporal-parietal junction (TPJ). Second, in a cross-sectional study (Richardson, Lisandrelli, Riobueno-Taylor, & Saxe, 2018), children ages 3–12 years watched a silent animated movie, Pixar's "Partly Cloudy", while undergoing fMRI; in adults, this movie elicits the attribution of false beliefs and other mental states and thus activates the theory-of-mind brain network. Following the fMRI scan, children received a battery of behavioral tasks, including traditional false-belief tasks. Analyses of children's developing theory-of-mind network provided no support for "the prediction of a robust discontinuity in response, associated with the transition from failure to success on explicit false-belief tasks. In the profiles of neural responses, we saw no major discontinuity when children begin to systematically pass false-belief tasks" (p. 8). Finally, in an fNIRS study, 7-month-old infants watched videos of three different transfer-of-location scenarios (Hyde et al., 2018): The agent faced away during the object's transfer from one container to another (false-belief scenario), she witnessed this transfer (true-belief scenario), or she faced away during the transfer but could infer it when she turned back because the containers were transparent (direct-perception scenario). In each scenario, activation in the TPJ was measured prior to the agent's search for the object. Like adults (Hyde et al, 2015), infants showed more TPJ activation during the false-belief scenario than during the true-belief and direct-perception scenarios, providing no support for claims of marked discontinuities in the development of false-belief understanding.

False beliefs about identity. Positive findings from various non-traditional tasks call into question the claim that infants are unable to represent false beliefs about identity (Buttelmann, Suhrke, & Buttelmann, 2015; Forgács et al., 2019; Scott & Baillargeon, 2009; Song & Baillargeon, 2008). For example, in one violation-of-expectation task (Scott & Baillargeon, 2009), 18-month-olds first received familiarization trials in which an agent sat behind a 1-piece toy penguin that did not come apart and a disassembled 2-piece penguin (Figure 1). Across trials, an experimenter placed the 1-piece penguin and the disassembled 2-piece penguin on platforms or in shallow containers; the side of each penguin (left or right) varied across trials. In each trial, the agent hid a small key in the bottom piece of the 2-piece penguin and then assembled it; once assembled, the 2-piece penguin looked identical to the 1-piece penguin. In the test trials, the agent was initially absent; the experimenter assembled the 2-piece penguin, placed it under a transparent cover, and then placed the 1-piece penguin under an opaque cover. The agent returned with her key, reached for one of the covers, and then paused. Infants looked significantly longer when the agent reached for the opaque as opposed to the transparent cover (this effect reversed when the agent witnessed the experimenter's actions). These results indicated that infants expected the agent (a) to mistake the penguin visible under the transparent cover for the 1-piece penguin (because the 2-piece penguin was always disassembled at the start of the familiarization trials) and hence (b) to falsely conclude that the disassembled 2-piece penguin was hidden under the opaque cover (because both penguins were always present in the familiarization trials). These results thus demonstrate that infants can reason about the actions of an agent who mistakes one object (the assembled 2-piece penguin) for a different object (the 1-piece penguin).

Interlocking mental states. Positive findings from non-traditional tasks also cast doubt on the claim that infants are unable to represent multiple interlocking mental states that interact causally. For instance, to succeed in the task just described, infants had to represent the agent's intention of hiding her key in the 2-piece penguin, her mistaken belief that the penguin under the transparent cover was the 1-piece penguin, and her ensuing mistaken inference that the disassembled 2-piece penguin must be hidden under the opaque cover. Additional evidence comes from a study that explored 17-month-old infants' understanding of how an agent might go about implanting a false belief about the identity of an object in another agent (Scott, Richman, & Baillargeon, 2015). Infants were tested in a violation-of-expectation task in which a thief attempted to secretly steal a desirable rattling toy during its owner's absence by substituting a less desirable silent toy. Infants first watched a series of rattling-toy and silent-toy familiarization trials (Figure 2). In each rattling-toy trial, the owner entered with a different toy on a tray; she shook the toy, which rattled, placed it back on the tray and then left. In her absence, the thief picked up the toy, shook it, and then replaced it on the tray. When the owner returned, she stored the rattling toy in a treasure box. The silent-toy trials were similar except that the toy was silent, the thief did not play with it, and the owner discarded it in a trashcan. In the test trial, for the first time, the owner brought in a rattling test toy that was visually identical to a silent toy she had previously discarded in the trashcan. As before, she shook the toy and then left. In her absence, the thief picked up the rattling toy, peered into the trashcan, and selected either the matching silent toy (matching trial) or a non-matching silent toy (non-matching trial). The thief placed the silent toy on the owner's tray, hid the rattling test toy in a pocket, and then paused. Infants looked significantly longer if shown the non-matching as opposed to the matching trial, suggesting that they realized that (a) the thief sought to steal the rattling test toy without the owner's knowledge by substituting a discarded silent toy for the rattling test toy, and (b) the thief could achieve this deceptive goal only by substituting the matching silent toy. Further results indicated that infants expected the returning owner to be deceived by this substitution and to store the matching silent toy in her treasure box. These results provide evidence that infants can reason about multiple interlocking mental states: To succeed at the task, infants had to represent the thief's deceptive goal of secretly stealing the rattling test toy by implanting in the owner a false belief about the identity of the toy on the tray, and they had to reason that the thief could achieve this goal only by substituting a silent toy that was visually identical to the rattling toy—otherwise the owner would detect the substitution as soon as she returned and saw the silent toy.

Failure at traditional tasks. The central tenet of two-system accounts is that the early-developing system cannot support success at traditional, elicited-prediction tasks—only the late-developing system can do so. According to one-system accounts, however, young children who have sufficient verbal ability to comprehend the test question (e.g., "Where will Sally look for her marble?") should be able to determine the correct answer to this question as long as processing difficulties are sufficiently reduced, and there is growing evidence for this prediction. First, toddlers demonstrate false-belief understanding when the test question is addressed to a third party (Barrett et al., 2013; Garnham & Ruffman, 2001; He, Bolz, & Baillargeon, 2012; Scott et al., 2012). In a non-traditional violation-of-expectation task, for example, 2.5-year-olds found it unexpected when an adult who was asked where Sally would look for her toy answered incorrectly by pointing to the toy's current location, as opposed to the location corresponding to Sally's false belief (Scott et al., 2012). In line with the expanded-processing-demands view, because the test question was not directed at the toddlers, no response-generation process was initiated and no prepotent incorrect response had to be inhibited, enabling toddlers to correctly determine how the adult should answer the question.

Second, there is also evidence that when processing demands are appropriately reduced, toddlers can even correctly answer the test question themselves (Setoh, Scott, et al., 2016; for a replication, see Grosso, Schuwerk, Kaltefleiter, & Sodian, 2019). In a modified traditional task, toddlers heard a transfer-of-location story accompanied by a picture book: Emma found an apple in one of two containers, moved it to the other container, and then went outside to play with her ball; in her absence, her brother Ethan found the apple and took it away. Emma then returned to look for her apple. In the test trial, children were shown pictures of the two containers and were asked the test question, "Where will Emma look for her apple?" The task included two modifications designed to reduce processing demands. One modification was that Emma's brother took the apple away to an undisclosed location instead of moving it to the other container. This modification built on findings that children age 3.5 years and older typically perform better at traditional tasks if the test object is absent from the scene (Bartsch, 1996; Wimmer & Perner, 1983), presumably because children then require less inhibitory control to suppress their knowledge of the object's current location when asked the test question. The other modification was that during the story, toddlers were asked two practice questions designed to reduce the response-generation demands of the test question. In one practice question, toddlers saw an apple and a banana and were asked, "Where is Emma's apple?"; in the other question, they saw a ball and a frisbee and were asked, "Where is Emma's ball?" These questions thus gave children practice interpreting a "where" question and producing a response by pointing to one of two pictures. With these two modifications, toddlers performed above chance in the test trial, pointing to the container Emma falsely believed held her apple. Additional experiments indicated that toddlers performed at chance if they received one or fewer practice trials, or if they received two practice trials that differed in form from the test trial, rendering them less effective at reducing response-generation demands.

Together, the results reviewed in this section call into question the claims by two-system theorists that early psychological reasoning has a distinct neurological substrate and cannot support the attribution of false beliefs about identity, the representation of interlocking mental states, or success at traditional false-belief tasks. To the contrary, these results suggest that young children's psychological reasoning is qualitatively similar to that of older children and adults, thereby providing evidence for one-system accounts.

However, there remains one signature limit proposed by two-system accounts that is not addressed by the preceding results: cognitive encapsulation. Thus, proponents of these accounts could argue that a critical distinction might still exist between early and more mature forms of psychological reasoning: It could be that psychological reasoning is at first largely encapsulated from the rest of cognition and, as such, tends to be inflexible, with sharp limits on children's ability to integrate information from other cognitive processes when reasoning about agents' actions. We evaluate this suggestion in the next section.

B. Cognitive Encapsulation

Is early psychological reasoning largely cognitively encapsulated, as two-system accounts suggest? To our knowledge, this question has received little discussion so far. In this section, we address this issue by examining two well-studied examples from the developmental literature: early reasoning about preferences and about false beliefs. These examples paint a picture of early psychological reasoning more consistent with one-system accounts, in which cognitive processes and world knowledge are readily integrated, in a flexible and context-sensitive manner, into children's psychological reasoning.

1. Attributing Preferences to Others

In Woodward's (1998) seminal two-object task, 5-, 6-, and 9-month-olds first received familiarization trials in which an agent faced two different objects, object-A and object-B; in each trial, she grasped object-A and paused. In the test trials, the objects' locations were swapped, and the agent reached either for object-A in its new position (old-object event) or for object-B (new-object event). At all ages, infants looked longer at the new-object than at the old-object event. This result, which was widely replicated and extended to infants as young as 3 months of age (Johnson et al., 2007; Luo, 2011; Martin, Onishi, & Vouloumanos, 2012; Spaepen & Spelke, 2007), suggests that (a) infants interpreted the agent's unvarying choice of object-A over object-B during the familiarization trials as signaling a preference for object-A; (b) they expected her to continue acting on this preference during the test trials, in accordance with the rationality principle; and (c) they therefore looked longer when she violated this expectation by reaching for object-B instead. Support for this interpretation came in part from one-object tasks that were identical to Woodward's two-object task except that only object-A was present during the familiarization trials (Bíró, Verschoor, & Coenen, 2011; Choi, Mou, & Luo, 2018; Luo & Baillargeon, 2005; Song, Baillargeon, & Fisher, 2014). Infants now looked equally at the old- and new-object events, suggesting that when no choice information was available in the familiarization trials, infants had no basis for attributing a preference to the agent and therefore held no expectations about which object she would select in the test trials.

The preceding interpretation of the two-object and one-object tasks is broadly consistent with the capabilities of the early-developing psychological-reasoning system, as described by two-system theorists. The same is true of several variants of these tasks. For example, the early-developing system could explain why infants fail to attribute to an agent a preference for object-A over object-B when the agent cannot see object-B in the familiarization trials (Kim & Song, 2015; Luo & Baillargeon, 2007), or when the agent is replaced by a different agent in the test trials (Buresh & Woodward, 2007; Henderson & Woodward, 2012). To the extent that the early-developing system can track simple epistemic states (e.g., what objects the agent can see) and motivational states (e.g., what objects the agent prefers) and views these as agent-specific, it is sufficient to explain these findings. However, this is not true of other findings from two-object tasks that require the integration of information from other cognitive processes and are thus inconsistent with a cognitively encapsulated system. We highlight five such findings here.

Physical reasoning. In a recent modified two-object task, researchers examined whether 5-month-old infants would bring to bear their physical knowledge that wide objects cannot fit into narrow containers (Wang, Baillargeon, & Brueckner, 2004) when reasoning about an agent searching for a preferred wide toy (Ting, He, & Baillargeon, 2019). In each familiarization trial, the agent first saw an experimenter shake a wide toy that rattled and a narrow toy that did not (the side of each toy varied across trials); after the experimenter left, the agent consistently reached for and shook the wide toy, suggesting that she preferred it. In the test trial, the agent was initially absent, and the experimenter hid the wide toy in a wide box and the narrow toy in a narrow box; both toys had identical handles that protruded from the boxes. Next, the agent returned and reached for either the wide or the narrow box. Infants looked significantly longer if they saw the agent reach for the narrow as opposed to the wide box (this effect was eliminated when both boxes were wide, giving the agent no cue to the wide toy's location). These results suggest that infants were able to integrate their physical knowledge into their reasoning about the agent's actions: They expected her to share their knowledge that wide objects can fit only into wide containers and to search for her preferred object accordingly.

Interpreting a novel label. By 6 months of age, infants understand the communicative function of speech and perceive words as linguistic conventions that are shared by members of the same speech community (Martin et al., 2012; Vouloumanos, Martin, & Onishi, 2014). This means that if agent-1 requests her preferred object from agent-2 using a novel label, infants expect agent-2 to understand this label and hence to know which object is being referred to. In one experiment, 6-month-olds first received familiarization trials in which agent-1 repeatedly selected object-A over object-B (Vouloumanos et al., 2014). In the test trial, agent-1 could no longer reach the objects (she peered at them through a small window) and was joined by agent-2, who could reach both objects. Agent-1 looked at agent-2 and said, "Koba, koba!" Agent-2 then picked up either object-A (old-object event) or object-B (new-object event) and offered it to agent-1. Infants looked significantly longer if shown the new-object as opposed to the old-object event (this effect was eliminated if agent-1 coughed instead). These results suggest that infants interpreted agent-1's novel word as referring to her preferred object, and they expected agent-2 to share this linguistic knowledge and hence to know which object agent-1 was requesting.

Interpreting a change in label. Infants understand from a young age that words can serve as labels for objects (Bergelson & Swingley, 2012), and this has consequences for two-object tasks in which the agent utters a novel word in each trial prior to reaching for an object (Jin & Song, 2017; Song et al., 2014). In one experiment (Jin & Song, 2017), 12- and 14-month-olds first received familiarization trials in which the agent said a novel word (e.g., "Modi!") prior to selecting object-A as opposed to object-B. In the test trials, the agent said a different novel word (e.g., “Papu!”) and then grasped either object-A (old-object event) or object-B (new-object event). At both ages, infants looked significantly longer at the old-object than at the new-object event, the opposite pattern from that usually found. These results suggest that infants did not interpret the agent’s actions in the familiarization trials as demonstrating a preference for object-A; rather, they assumed that the label the agent produced signaled which object she would reach for, and they therefore interpreted the change in what she said in the test trials as signaling a change in which object she would grasp next.

Categorizing objects. Beginning around their first birthday (perhaps due to advances in language acquisition), infants begin to spontaneously encode objects’ basic-level categories (e.g., duck, ball, truck, or doll), and this has consequences for two-object tasks (Liu & Sun, 2018; Spaepen & Spelke, 2007). After watching familiarization trials in which an agent consistently chose object-A over object-B (Spaepen & Spelke, 2007), 12-month-olds expected the agent to reach for object-A again in the test trials if the two objects were from different basic-level categories (e.g., a truck and a doll), but not if both objects were from the same category (e.g., two trucks that were perceptually distinct). These results suggest that infants interpreted the agent’s repeated selection of object-A as signaling a preference for an object category (e.g., trucks), rather than a preference for a particular object (e.g., this truck). Thus, when the two objects were from the same category (e.g., trucks), infants had no expectation about which object the agent would select. Conversely, when the objects belonged to two different object categories and were replaced in the test trials with novel exemplars of these categories (e.g., a novel truck and a novel doll), infants still expected the agent to reach for the exemplar from the preferred category (i.e., the novel truck). These results suggest that at least by 12 months of age, infants integrate information about object categories when reasoning about agents’ preferences.

Interpreting ostensive cues. According to natural-pedagogy theory, infants interpret ostensive cues as signaling "teaching" episodes intended to communicate general information about objects or events (Csibra & Gergely, 2009). Ostensive cues have been shown to affect responses in two-object tasks in which agents direct different emotional displays at object-A and object-B. In one experiment (Egyed, Király, & Gergely, 2013), 18-month-olds first saw agent-1 express positive affect toward object-A and negative affect toward object-B. When no ostensive cues were used, infants assumed that agent-1 preferred object-A, and they were significantly more likely to select it in the test trial when she requested one of the objects ("Give me one of them!"). Moreover, consistent with prior findings, infants chose randomly between the two objects if agent-1 was replaced by agent-2 in the test trial. However, when agent-1's emotional displays were accompanied by ostensive cues (i.e., agent-1 called the infant's name, said, "Look!", and alternated her gaze between the infant and the object), infants now gave object-A to agent-2 in the test trial. This last result suggests that infants (a) assumed that agent-1's ostensive displays were intended to teach them that object-A was pleasing but object-B was not, and (b) expected agent-2 to share this knowledge and hence to also prefer object-A.

2. Attributing False Beliefs to Others

As was discussed earlier, over 30 reports have provided positive evidence of false-belief understanding in the first three years of life. These reports have used many different methods and have involved many different scenarios, with mistaken agents holding different false beliefs and producing different responses ranging from physical actions and emotional reactions to verbal responses (Scott & Baillargeon, 2017). In this section, we focus on a subset of these reports to illustrate that, far from being encapsulated, early false-belief understanding is flexible, context-sensitive, and well integrated with other cognitive processes.

Physical reasoning. In a recent violation-of-expectation task, researchers examined whether 10-month-olds would bring to bear their physical knowledge that tall objects cannot be hidden in short containers (Hespos & Baillargeon, 2001) when reasoning about an agent who was playing a hide-and-seek game and searching for an object she mistakenly believed to be tall (Bian, He, & Baillargeon, 2019). In each familiarization trial, an agent sat at a window in the back wall of a puppet-stage apparatus behind a tall toy dog and two short containers. While an experimenter watched from a side window, the agent played with the dog briefly, put it down, and then hid herself by lifting a large cloth that filled her window ("I'm ready!"). Next, the experimenter placed the dog in one of the containers and then signaled the agent to return ("Ok!"). At that point, the agent lowered her cloth, grasped the dog's head, and paused. Across trials, different containers were used, and the side of the dog (left or right) was varied. The test trial was identical, with two exceptions: The containers were a tall and a short box, each closed with a lid, and the experimenter shortened the dog (its body could be collapsed by pressing firmly on its head) and hid it in the short box. When the agent returned, she grasped the lid of either the tall or the short box and then paused. Infants looked significantly longer if they saw the agent reach for the short as opposed to the tall box. These results suggest that infants expected the agent (a) to falsely believe that the dog was still tall and hence (b) to falsely conclude that it must be hidden in the tall box (this effect was eliminated if the agent saw the dog being shortened before hiding herself, as the short dog could then be hidden in either box). Infants were thus able to integrate their physical knowledge into their reasoning about the agent's actions: They expected her to share their knowledge that tall objects can be hidden only in tall containers and to search accordingly.

Detecting semantic incongruity. By the second year of life, infants' burgeoning word knowledge enables them to detect a violation when a spoken label is incongruent with an agent's false belief about an object's identity. In a semantic-incongruity task measuring event-related potentials (ERPs; Forgács et al., 2019), 14-month-olds faced an agent who sat behind an occluder. A familiar object (e.g., an apple) was placed in front of the occluder, where it was visible to the infants but not the agent. The occluder was then lowered and the agent inspected the object. Next, the occluder was raised, the agent turned away, and the original object was replaced with a different familiar object (e.g., a car). The agent then turned back, and the occluder either remained raised (false-belief trials) or was lowered to reveal the new object (true-belief trials). Infants then heard a label that matched the current object ("car"). Analyses focused on the N400, an ERP component that indexes semantic processing: Adult listeners typically produce larger N400s to semantically incongruous statements (Kutas & Federmeier, 2011). In line with these results, infants exhibited significantly larger N400s to the label during the false-belief trials than during the true-belief trials, suggesting that they detected a violation when the label (though accurate from their perspective) was incongruous with the agent's false belief about the object's identity.

Sociomoral reasoning. From a young age, infants expect agents to approach someone with a good moral character, but to avoid someone with a bad moral character (Hamlin et al., 2007). This sociomoral knowledge has been shown to be integrated with infants' false-belief reasoning. In a violation-of-expectation task (Choi & Luo, 2015), 15-month-olds first saw agent-1 and agent-2 interact positively. In the next trial, while agent-1 was absent, agent-2 deliberately knocked down another agent, agent-3. In the test trials, agent-1 either ignored agent-2 (alone event) or continued to interact positively with her as before (together event). Infants looked significantly longer at the alone than at the together event (this effect reversed if agent-1 witnessed agent-2's harmful actions toward agent-3). Infants thus integrated their sociomoral knowledge into their reasoning about agent-1's actions: They expected her to continue interacting with agent-2 as long as she falsely believed in agent-2's good moral character.

Inferences based on general knowledge. In the false-belief tasks described thus far, the agent's false belief was based on information that was acquired during the experimental situation and became invalid (e.g., the toy was moved from box-A to box-B in the agent's absence). Other tasks have revealed that infants can also attribute false beliefs based on general knowledge about the world that happens not to hold true in the experimental situation (He, Bolz, & Baillargeon, 2011; Scott, Baillargeon, Song, & Leslie, 2010). One violation-of-expectation task (Scott et al., 2010) built on prior findings that the greater the visual similarity between two objects, the greater the likelihood that 18-month-olds will generalize a non-obvious property from one object to the other (Welder & Graham, 2001). In the task, 18-month-olds first received a familiarization trial. An experimenter sat at a side window in a puppet-stage apparatus and faced a red toy with silver stars. Two additional toys stood at the rear of the apparatus: One was visually identical to the experimenter's toy (red test toy) and the other was green with yellow stripes (green test toy). The experimenter shook each toy in turn, thereby revealing that her toy and the green test toy rattled when shaken but the red test toy did not. In the test trial, an agent now sat behind the two test toys. The experimenter shook her toy and said, "Can you do it?" The agent then grasped either the red or green test toy and paused. Infants looked significantly longer if they saw the agent select the green as opposed to the red test toy (this effect reversed if she witnessed the initial shaking of the toys). These results suggest that infants expected the agent (a) to share their knowledge that visually identical objects are more likely to share non-obvious properties than are non-identical objects and hence (b) to falsely expect the red test toy to rattle like the experimenter's toy.

Updating an agent's false belief. Because infants realize that language conveys information between communicative partners, they understand that a mistaken agent's false belief can be updated or corrected by an informed agent's appropriate utterance. In a violation-of-expectation task (Song, Onishi, Baillargeon, & Fisher, 2008), 18-month-olds first saw an agent place a ball in a box. Next, while the agent was absent, an experimenter moved the ball to a nearby cup. The agent then returned, and the experimenter produced an utterance about the ball's location that was either informative ("The ball is in the cup!") or uninformative ("I like the cup!"). In the test trial, the agent was alone and reached for either the cup or the box and then paused. Infants who heard the informative utterance expected the agent to reach for the cup, suggesting that they considered this utterance sufficient to update the agent about where to find the ball. In contrast, infants who heard the uninformative utterance expected the agent to reach for the box, suggesting that they deemed the utterance insufficient to update the agent and hence expected her to continue believing that the ball was in the box.

Determining the referent of a novel label. By 17 months of age, infants consider an agent's beliefs about a scene when determining the likely referent of the agent's verbal utterances (Southgate et al., 2010). In a referential-communication task, 17-month-olds first saw an agent place one novel object in box-A and a second novel object in box-B; the agent then left the room, and an experimenter swapped the locations of the objects. When the agent returned, she pointed to box-A, said it contained a "sefo", and asked the child, "Can you get the sefo for me?" Infants were significantly more likely to approach box-B, the box the agent had not pointed to (this effect reversed when the agent witnessed the swap of the objects' locations). These results suggest that infants recognized that the agent held a false belief about the objects' locations and intended to refer to the object in box-B when labeling the sefo.

Revising a belief that was incorrectly computed. The ability to retrieve and reason about past events appears to improve markedly around age 3 (Hayne, 2004), and this enables children to retrieve and revise a belief they incorrectly attributed to an agent. In a modified referential-communication task (Király, Oláh, Csibra, & Kovács, 2018), 36-month-olds first saw an agent place two novel objects into two boxes; she then put on a pair of sunglasses and faced the boxes while the experimenter silently swapped the objects' locations. Next, the agent took off the sunglasses, put them in front of the child, and left the room with the experimenter. In their absence, the parent helped the child inspect the sunglasses and discover that they were completely opaque. The agent then returned, pointed to box-A, indicated that it contained a sefo, and asked the child to give it to her. Children were significantly more likely to approach box-B as opposed to box-A (this effect reversed when the sunglasses were transparent). These results suggest that children (a) retrieved the memory that the agent had been wearing the sunglasses during the swap, (b) inferred that she did not see the swap and must falsely believe the objects were still in their original locations, and (c) concluded that she intended to refer to the object in box-B when labeling the sefo. Unlike 36-month-olds, 18-month-olds expected the agent to know the objects' current locations whether the sunglasses were transparent or opaque. Thus, consistent with prior evidence on the development of episodic memory, only the 36-month-olds were able to retrospectively revise the belief they had attributed to the agent after learning about the sunglasses' properties. This ability is, incidentally, the same one that was illustrated at the start of this chapter: We first attributed to Anita the true belief that the woman down the aisle was her trainee, but later inferred that Anita had in fact mistaken her trainee's identical twin for her trainee.

3. Interim Summary

Together, the 12 sets of findings described in the preceding sections on preference and false-belief tasks suggest that far from being cognitively encapsulated, early psychological reasoning is flexible, context-sensitive, and closely entwined with the rest of cognition. When reasoning about an agent's actions, infants and toddlers are able to integrate information from many different cognitive processes, including physical reasoning, sociomoral reasoning, categorization, language comprehension, general knowledge, and episodic memory. Moreover, as new knowledge is acquired and as new cognitive abilities emerge, children appear to immediately start integrating them into their psychological reasoning (Carruthers, 2018). These findings, together with those reviewed earlier that call into question other signature limits proposed by two-system accounts, are all deeply problematic for these accounts. Instead, these various findings provide strong support for one-system accounts.

V. Future Directions

The evidence reviewed in this chapter is most consistent with the notion that a single mentalistic system guides human psychological reasoning, emerging early in infancy and gradually developing in sophistication over time. However, much remains to be understood about this development.

One important research direction will be to focus on more subtle facets of psychological reasoning. For example, there is extensive evidence that children become better over time at understanding that an agent's verbal statements or emotional displays do not always reflect what the agent truly thinks or feels, as in irony and sarcasm, white lies and other forms of benevolent social acting, and face-saving situations where one chooses to hide discomfiture from others (Baillargeon et al., 2013; Ma, Xu, Heyman, & Lee, 2011; Peterson, Wellman, & Slaughter, 2012; Wellman, Fang, & Peterson, 2011). Research efforts are getting under way to study at what age children first begin to understand such counterfactual displays.

Another important research direction, which is currently receiving a great deal of attention, is to better understand what factors drive performance in various psychological-reasoning tasks.

We discuss two sets of relevant findings below.

A. Non-Replications and Component Abilities

Over the past few years, there have been several reports of failed attempts at replicating positive findings with infants and toddlers in non-traditional false-belief tasks (Crivello & Poulin-Dubois, 2018; Dörrenberg, Rakoczy & Liszkowski, 2018; Kulke, Reiß, Krist, Rakoczy, 2018; Poulin-Dubois, Polonia, & Yott, 2013; Powell, Hobbs, Bardis, Carey, & Saxe, 2018; Priewasser, Rafetseder, Gargitter, & Perner, 2018; Yott & Poulin-Dubois, 2016). A few researchers have taken these non-replications to call into question the very existence of an early capacity for false-belief understanding (Kulke et al., 2018; Poulin-Dubois et al., 2018; Powell et al., 2018). Such a conclusion is unlikely, however: As we saw earlier, over 30 reports, using 11 different behavioral and neural methods, have yielded positive evidence of early false-belief understanding in non-traditional tasks. To paraphrase Carruthers (2018), this represents too many studies, using too many variations in methods and stimuli, for such a conclusion to be plausible. A more constructive approach to these non-replications, we believe, is to examine what they reveal about the component abilities that contribute to success or failure at these tasks.

To draw an analogy, we saw earlier that 2.5-year-old toddlers succeed at a traditional task using a transfer-of-location scenario when (a) inhibitory-control demands are reduced by removing the test object from the scene and (b) response-generation demands are also reduced by providing two practice questions similar in structure to the test question (Setoh, Scott, et al., 2016). In contrast, toddlers fail when either of these modifications is altered. Together, this mix of positive and negative findings helps bring to light some of the component abilities that allow toddlers to correctly answer questions about others' false beliefs. In the same vein, we believe that negative findings in non-traditional tasks help shed light on some of the component abilities that contribute to success at these tasks.

In some non-replications, critical components were revealed by procedural changes that led to negative results (Crivello & Poulin-Dubois, 2018; Powell et al., 2018; Priewasser, et al., 2018; Poulin-Dubois et al., 2013). To illustrate, consider two attempts at replicating the violation-of-expectation task of Onishi and Baillargeon (2005). In this task, all infants received a belief-induction trial that was separate from the test trial. For instance, infants in one condition received a belief-induction trial in which the toy moved in the agent's absence from box-A to box-B; this event was followed by a paused scene that ended when infants looked away and the trial ended. Because the toy moved for the first time in a self-propelled manner, this paused scene allowed infants time to process this new information and to work out its implications for the agent's actions. In their close replication of Onishi and Baillargeon (2005), Yott and Poulin-Dubois (2012) also used a belief-induction trial and replicated the original positive results. In contrast, Powell et al. (2018) did not use a separate belief-induction trial and obtained negative results. In the test trial, while the agent was briefly out of sight, the toy either moved from one box to the other on its own or was moved by a bystander; in either task, the agent reached for a box immediately after she returned. Together, these results suggest that in order to succeed at a violation-of-expectation false-belief task, infants must have sufficient time to process the relevant information and to form an expectation about what the agent will do before she initiates her actions (see also Schulze & Buttelmann, 2017).

In other non-replications, negative findings are less easily traced to procedural changes but seem to arise from more diffuse factors affecting participants' motivation. This is particularly true for the non-replications of the anticipatory-looking task Southgate et al. (2007) originally used with 25-month-old toddlers and later extended to infants, older children, and adults (Senju et al., 2010; Senju, Southgate, Snape, Leonard, & Csibra, 2011; Senju, Southgate, White, & Frith; 2009). In this task, participants must anticipate where an agent who holds a false belief about an object's location will search for it. Failed replications have been reported with toddlers, older children, and adults (Dörrenberg et al., 2018; Grosse Wiesmann, Friederici, Disla, Steinbeis, & Singer, 2018; Kulke et al., 2018; Kulke, von Duhn, Schneider, & Rakoczy, 2018; Schuwerk, Priewasser, Sodian, & Perner, 2018). When commenting on these findings, Southgate made two observations (Baillargeon, Buttelmann, & Southgate, 2018). First, the non-replications showed less correct anticipation in the true-belief familiarization trials than did the original studies, raising questions about participants' motivation to predict the agent's action. Second, test results across the non-replications showed no consistent pattern, with some studies reporting above-chance performance in one false-belief condition and below-chance performance in another, and other studies reporting at-chance performance across conditions. Together, these observations suggest that an anticipatory-looking false-belief task may yield positive findings only when everything about the task conspires to make participants highly engaged by the agent's actions so that they are focused on predicting what she will do next. At present, there is a multi-lab project under way (the "ManyBabies-2" project; e.g., Kampis & Hamlin, 2019) that is attempting to determine under what conditions infants are likely to robustly anticipate agents' actions.

B. Individual Differences

As the previous discussion illustrates, the capacity to represent mental states does not necessarily guarantee successful performance in a psychological-reasoning task. Rather, success can be influenced by properties of the task, as we just saw, as well as by differences across individuals. Decades of research with older children and adults has identified a number of factors that contribute to individual differences in performance on psychological-reasoning tasks. These include cognitive abilities (Carlson & Moses, 2001; Devine & Hughes, 2014; Duh et al., 2016; Milligan et al., 2007) as well as social factors such as exposure to conversations about mental states (Adrián, Clemente, & Villanueva, 2007; Ensor & Hughes, 2008; Ruffman et al., 2002; Slaughter, Peterson, & McKintosh, 2007), socioeconomic status (Cutting & Dunn, 1999; Devine & Hughes, 2018; Holmes, Black, & Miller, 1996; Hughes et al., 2000), and cultural background (Lecce & Hughes, 2010; Liu et al., 2008; Mayer & Träuble, 2013; Naito & Koyoma, 2006). To date, few studies have examined individual variation in infants' and toddlers' performance on the types of psychological-reasoning tasks that we have reviewed here. Thus, relatively little is known about individual differences in early psychological reasoning and the factors that might be responsible for such differences.

Researchers have recently begun to address these questions by examining whether the factors that predict psychological reasoning in preschool children also predict individual differences in toddlers' performance on non-traditional false-belief tasks. In one study, individual differences in verbal ability predicted 3-year-olds' performance on a preferential-looking task involving high verbal demands (Scott & Roby, 2015). Another study demonstrated that 2.5-year-olds' exposure to mental-state terms (e.g., *think*, *know*) in conversations with their parents predicted their performance on an anticipatory-looking task (Roby & Scott, 2018). Together, these findings suggest continuity in the factors that influence psychological reasoning across early childhood. These findings also raise the possibility that at least some of the non-replications of non-traditional tasks stem from individual variability: If some individuals are highly motivated to anticipate an agent's behavior whereas others are less so, this could produce chance performance at the group level. Exploring the factors that predict individual differences in early psychological reasoning, as well as the mechanisms underlying these relationships, thus presents a promising avenue for future research.

VI. Conclusion

How sophisticated is infants' theory of mind? The evidence reviewed in this chapter suggests two broad conclusions. First, infants' psychological reasoning is driven by a single mentalistic system that is qualitatively similar to that of older children and adults. From a young age, this system enables infants (a) to infer agents' motivational, epistemic, and counterfactual mental states; (b) to attribute false beliefs with a wide range of propositional content, including false beliefs about identity; (c) to represent, under some conditions at least, complex arrays of interlocking mental states that interact causally; (d) to integrate world knowledge and information from other cognitive processes into their reasoning about agents' likely mental states; and finally (e) to use agents' mental states, together with a principle of rationality, to predict, interpret, and evaluate agents' actions and to guide their own actions towards agents.

Second, although infants' psychological reasoning system appears to be broadly similar to that of older children and adults, it is nevertheless considerably less sophisticated. The evidence we have reviewed points to at least three important differences between early and more mature psychological reasoning. One difference is that with age, children become better able to cope with traditional and non-traditional false-belief tasks with high processing demands. Gradual improvements in cognitive capacities such as inhibitory control, working memory, and verbal ability help children better handle these demands, as does everyday practice at conversing about mental states with parents, siblings, and peers. Another difference is that when attempting to infer an agent's mental states, infants and toddlers cannot take into account world knowledge they have not yet acquired or information from cognitive processes (such as episodic memory) they are still developing. Finally, children become more skilled over time at understanding subtle facets of psychological reasoning, such as when agents offer verbal and emotional communications that do not convey what they truly think or feel.

Footnote

1. Positive findings have also been obtained with children between 3 and 4 years of age in a wide range of modified traditional and non-traditional tasks with reduced processing difficulties (Hansen, 2010; Rhodes & Brandone, 2014; Rubio-Fernandez & Geurts, 2013; Yazdi, German, Defeyter, & Siegal, 2006). However, proponents of conceptual-change accounts have argued that such results could simply be due to transitional children beginning to show the more mature form of psychological reasoning that typically emerges around age 4 (Gopnik & Wellman, 1994; Wellman et al., 2001). For this reason, we focus here on children age 36 months and younger.

References

Adrián, J. E., Clemente, R. A., & Villanueva, L. (2007). Mothers' use of cognitive state verbs in picture-book reading and the development of children's understanding of mind: A longitudinal study. *Child Development*, *78,* 1052-1067.

Antilici, F. (2019). The processing-time account: Why young children fail direct false-belief tests despite possessing false-belief understanding. Manuscript under review.

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

Baillargeon, R., Buttelmann, D., & Southgate, V. (2018). Invited Commentary: Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations. *Cognitive Development, 46*, 112-124.

Baillargeon, R., He, Z., Setoh, P., Scott, R. M., Sloane, S., & Yang, D. Y. J. (2013). False-belief understanding and why it matters. In R. Mahzarin & S.A. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 88-95). Oxford, England: Oxford University Press.

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., Sloane, S., Setoh, P., Jin, K., Wu, D., & 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.

Bardi, L., Desmet, C., Nijhof, A., Wiersema, J. R., & Brass, M. (2017). Brain activation for spontaneous and explicit false belief tasks overlaps: New fMRI evidence on belief processing and violation of expectation. *Social Cognitive and Affective Neuroscience*, *12*, 391-400.

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

Barrett, H. C., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., ... & Laurence, S. (2013). Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society of London B: Biological Sciences*, *280,* 20122654.

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

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*, 611-617.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*, 3253-3258.

Bian, L., He, Z., & Baillargeon, R. (2019). False-belief reasoning in the first year of life. Manuscript in preparation.

Bíro, S., Verschoor, S., & Coenen, L. (2011). Evidence for a unitary goal concept in 12-month-old infants. *Developmental Science, 14*, 1255-1260.

Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, *104*, 287-314.

Butler, A. G. (2013). Exploring the role of social reasoning and self-efficacy in the mathematics problem-solving performance of lower-and higher-income children. *Journal of Educational Research and Practice*, *3,* 93-119.

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

Buttelmann, F., Suhrke, J., & Buttelmann, D. (2015). What you get is what you believe: Eighteen-month-olds demonstrate belief understanding in an unexpected-identity task. *Journal of Experimental Child Psychology*, *131*, 94-103.

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

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, *9*, 55-72.

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

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

Carruthers, P. (2016). Two systems for mindreading? *Review of Philosophy and Psychology*, *7*, 141-162.

Carruthers, P. (2018). Young children flexibly attribute mental states to others. *Proceedings of the National Academy of Sciences*, *115*, 11351-11353.

Choi, Y. J., & Luo, Y. (2015). 13-month-olds' understanding of social interactions. *Psychological Science*, *26*, 274-283.

Choi, Y. J., Mou, Y., & Luo, Y. (2018). How do 3-month-old infants attribute preferences to a human agent? *Journal of Experimental Child Psychology*, *172*, 96-106.

Cowell, J. M., Lee, K., Malcolm-Smith, S., Selcuk, B., Zhou, X., & Decety, J. (2017). The development of generosity and moral cognition across five cultures. *Developmental*

Science, 20, e12403

Crivello, C., & Poulin-Dubois, D. (2018). Infants' false belief understanding: A non-replication of the helping task. *Cognitive Development, 46,* 51-57.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition, 107,* 705-717.

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27*, 111-133.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*, 148-153.

Csibra, G., Gergely, G., Bíró, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of 'pure reason' in infancy. *Cognition*, *72*, 237-267.

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development,* 70, 853-865.

de Villiers, J. & de Villiers, P. (2003). Language for thought: Coming to understand false beliefs. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 335-384). Harvard, MA: MIT Press.

Dennett, D. C. (1987). *The intentional stance.* Cambridge, MA: MIT Press.

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

Devine, R. T., & Hughes, C. (2018). Family correlates of false belief understanding in early childhood: A meta-analysis. *Child Development, 89*, 971-987.

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, *46,* 12-30.

Duh, S., Paik, J. H., Miller, P. H., Gluck, S. C., Li, H., & Himelfarb, I. (2016). Theory of mind and executive function in Chinese preschool children. *Developmental Psychology*, *52,* 582-591.

Dunfield, K. A., & Kuhlmeier, V. A. (2010). Intention-mediated selective helping in infancy. *Psychological Science*, *21*, 523-527.

Egyed, K., Király, I., & Gergely, G. (2013). Communicating shared knowledge in infancy. *Psychological Science*, *24*, 1348-1353.

Ensor, R., & Hughes, C. (2008). Content or connectedness? Mother-child talk and early social understanding. *Child Development*, *79*, 201-216.

Forgács, B., Parise, E., Csibra, G., Gergely, G., Jacquey, L., & Gervain, J. (2019). Fourteen-month-old infants track the language comprehension of communicative partners. *Developmental Science, 22*, e12751.

Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: Is anticipatory looking really related to understanding or belief? *Developmental Science*, *4*, 94-100.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences*, *7*, 287-292.

Gergely, G., Nádasdy, Z., Csibra G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165–193.

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.

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New

York, NY: Cambridge University Press.

Grosse Wiesmann, C. G., Friederici, A. D., Disla, D., Steinbeis, N., & Singer, T. (2018). Longitudinal evidence for 4-year-olds' but not 2-and 3-year-olds' false belief-related action anticipation. *Cognitive Development*, *46*, 58-68.

Grosso, S. S., Schuwerk, T., Kaltefleiter, L. J., & Sodian, B. (2019). 33-month-old children succeed in a false-belief task with reduced processing demands: A replication of Setoh et al. (2016). *Infant Behavior and Development*, *54*, 151-155.

Hamlin, J. K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, *18*, 451-474.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*, 557.

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development, 26*, 30–39.

Hansen, M. B. (2010). If you know something, say something: Young children's problem with false beliefs. *Frontiers in Psychology*, *1*, 23.

Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental Review*, *24*, 33-73.

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.

He, Z., Bolz, M., & Baillargeon, R. (2012). 2.5-year-olds succeed at a verbal anticipatory-looking false-belief task. *British Journal of Developmental Psychology, 30*, 14-29.

Helming, K. A., Strickland, B., & Jacob, P. (2016). Solving the puzzle about early beliefascription. *Mind & Language, 31,* 438-469.

Henderson, A. M., & Woodward, A. L. (2012). Nine-month-old infants generalize object labels, but not object preferences across individuals. *Developmental Science, 15*, 641-652.

Hespos, S. J., & Baillargeon, R. (2001). Infants' knowledge about occlusion and containment events: A surprising discrepancy. *Psychological Science*, *12*, 141-147.

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

Hofmann, S. G., Doan, S. N., Sprungc, M., Wilson, A., Ebesutanie, C., Andrews, L. A. … & Harris, P. L. (2016). Training children's theory-of-mind: A meta-analysis of controlled studies. *Cognition, 150,* 200-212.

Holmes, H. A., Black, C., & Miller, S. A. (1996). A cross-task comparison of false belief understanding in a Head Start population. *Journal of Experimental Child Psychology*, *63*, 263-285.

Hughes, C., Adlam, A., Happé, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good test-retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *Journal of Child Psychology and Psychiatry, 41,* 483-490.

Hyde, D. C., Aparicio Betancourt, M., & Simon, C. E. (2015). Human temporal-parietal junction spontaneously tracks others' beliefs: A functional near-infrared spectroscopy study. *Human Brain Mapping, 36*, 4831-4846.

Hyde, D. C., Simon, C. E., Ting, F., & Nikolaeva, J. I. (2018). Functional organization of the temporal–parietal junction for theory of mind in preverbal infants: A near-infrared spectroscopy study. *Journal of Neuroscience, 38,* 4264-4274.

Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B. & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. *Developmental Psychology, 52,* 1192-1205.

Jin, K. S., Houston, J. L., Baillargeon, R., Groh, A. M., & Roisman, G. I. (2018). Young infants expect an unfamiliar adult to comfort a crying baby: Evidence from a standard violation-of-expectation task and a novel infant-triggered-video task. *Cognitive Psychology, 102*, 1-20.

Jin, K. S., & Song, H. J. (2017). You changed your mind! Infants interpret a change in word as signaling a change in an agent's goals. *Journal of Experimental Child Psychology*, *162*, 149-162.

Johnson, S. C., Shimizu, Y. A., & Ok, S. J. (2007). Actors and actions: The role of agent behavior in infants' attribution of goals. *Cognitive Development*, *22,* 310-322.

Kamewari, K., Kato, M., Kanda, T., Ishiguro, H., & Hiraki, K. (2005). Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion. *Cognitive Development*, *20*, 303-320.

Kampis, D., & Hamlin, K. (2019, March). *ManyBabies 2: Theory of mind in infancy*. Paper presented at the biennial conference of the Society for Research in Child Development, Baltimore, MD.

Kampis, D., Parise, E., Csibra, G., & Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proceedings of the Royal Society B: Biological Sciences*, *282*, 20151683.

Kim, E. Y., & Song, H. J. (2015). Six-month-olds actively predict others' goal-directed actions. *Cognitive Development*, *33,* 1-13.

Király, I., Oláh, K., Csibra, G., & Kovács, Á. M. (2018). Retrospective attribution of false beliefs in 3-year-old children. *Proceedings of the National Academy of Sciences*, *115,* 11477-

11482.

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

Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: Possible mechanisms. *Developmental Psychology*, *46*, 815.

Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. *Cognitive Development*, *46*, 97-111.

Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychological Science*, *29*, 888-900.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647.

Lecce, S., & Hughes, C. (2010). The Italian job?: Comparing theory of mind performance in British and Italian children. *British Journal of Developmental Psychology*, *28*, 747-766.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review*, *94*, 412-426.

Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119-148). New York, NY: Cambridge University Press.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in ‘theory of mind’. *Trends in Cognitive Sciences*, *8*, 528-533.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*, 732-739.

Liu, S., & Sun, R. (2018). Do great minds prefer alike? Thirteen-month-old infants generalize personal preferences across objects of like kind but not across people. *Frontiers in Psychology, 9*, 2636.

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: a meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, *44*, 523-531.

Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives, 10*, 184-189.

Low, J., Drummond, W., Walmsley, A., & Wang, B. (2014). Representing how rabbits quack and competitors act: Limits on preschoolers' efficient ability to track perspective. *Child Development*, *85*, 1519-1534.

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, *24*, 305-311.

Luo, Y. (2011). Three-month-old infants attribute goals to a non-human agent. *Developmental Science, 14*, 453-460.

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, *16*, 601-608.

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*, 489-512.

Ma, F., Xu, F., Heyman, G. D., & Lee, K. (2011). Chinese children's evaluations of white lies: Weighing the consequences for recipients. *Journal of Experimental Child Psychology*, *108*, 308-321.

Margoni, F., Baillargeon, R., & Surian, L. (2018). Infants distinguish between leaders and bullies. *Proceedings of the National Academy of Sciences*, *115*, E8835-E8843.

Martin, A., Onishi, K. H., & Vouloumanos, A. (2012). Understanding the abstract role of speech in communication at 12 months. *Cognition*, *123*, 50-60.

Mayer, A., & Träuble, B. E. (2013). Synchrony in the onset of mental state understanding across cultures? A study among children in Samoa. *International Journal of Behavioral Development*, *37*, 21-28.

McAlister, A., & Peterson, C. C. (2006). Mental playmates: Siblings, executive functioning, and theory of mind. *British Journal of Developmental Psychology*, *24,* 733–751.

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

Meristo, M., & Surian, L. (2013). Do infants detect indirect reciprocity? *Cognition*, *129*, 102-113.

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

Moll, H., Khalulyan, A., & Moffett, L. (2017). 2.5-year-olds express suspense when others approach reality with false expectations. *Child Development*, *88*, 114-122.

Naito, M., & Koyama, K. (2006). The development of false-belief understanding in Japanese children: Delay and difference? *International Journal of Behavioral Development*, *30*, 290-304.

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

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J. (2010). Who took the cog out of cognitive science? Mentalism in an era of anti-cognitivism. In P.A. Frensch & R. Schwarzer (Eds.), *Cognition and neuropsychology: International perspectives on psychological science* (Vol. 1, pp. 241-261). Hove, England: Psychology Press.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5,* 125-137.

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

Poulin-Dubois, D., Brooker, I., & Polonia, A. (2011). Infants prefer to imitate a reliable person. *Infant Behavior and Development*, *34*, 303-309.

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., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., ... & Perner, J. (2018). Do infants understand false beliefs? We don't know yet–A commentary on Baillargeon, Buttelmann, and Southgate's commentary. *Cognitive Development*, *48*, 302-315.

Powell, L. J., Hobbs, K., Bardis, A., Carey, S., & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development*, *46*, 40-50.

Priewasser, B., Rafetseder, E., Gargitter, C., & Perner, J. (2018). Helping as an early indicator of a theory of mind: Mentalism or Teleology? *Cognitive Development*, *46*, 69-78.

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications*, *9*, 1027.

Rhodes, M., & Brandone, A. C. (2014). Three-year-olds' theories of mind in actions and words. *Frontiers in Psychology*, *5*, 263.

Roby, E., & Scott, R. M. (2018). The relationship between parental mental-state language and 2.5-year-olds' performance on a nontraditional false-belief task. *Cognition*, *180*, 10-23.

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

Rubio-Fernández, P., & Geurts, B. (2016). Don't mention the marble! The role of attentional processes in false-belief tasks. *Review of Philosophy and Psychology*, *7*, 835-850.

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

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development*, *73*, 734-751.

Schulze, C., & Buttelmann, D. (2017, March). *Toddlers' updating of belief attribution through indirect communicative acts*. Paper presented at the International Convention of Psychological Science, Vienna, Austria.

Schuwerk, T., Priewasser, B., Sodian, B., & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: a replication attempt. *Royal Society Open Science*, *5*, 172273.

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

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

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, *21*, 237-249.

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., He, Z., Baillargeon, R., & Cummins, D. (2012). False-belief understanding in 2.5-year-olds: Evidence from two novel verbal spontaneous-response tasks. *Developmental Science, 15*, 181-193.

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., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., ... & Csibra, G. (2010). Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Development and Psychopathology*, *22*, 353-360.

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

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

Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences*, *113*, 13360-13365.

Slaughter, V., Peterson, C. C., & Mackintosh, E. (2007). Mind what mother says: Narrative input and theory of mind in typical children and those on the autism spectrum. *Child Development, 78*, 839-858.

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science*, *23*, 196-204.

Song, H. J., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology*, *44,* 1789-1795.

Song, H., Baillargeon, R., & Fisher, C. (2014). The development of infants' use of novel verbal information when reasoning about others' actions. *PLoS ONE, 9*, e92387.

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.

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, *130*, 1-10.

Spaepen, E., & Spelke, E. (2007). Will any doll do? 12-month-olds' reasoning about goal objects. *Cognitive Psychology*, *54*, 133-154.

Takagishi, H., Kameshima, S., Schug, J., Koizumi, M., & Yamagishi, T. (2010). Theory of mind enhances preference for fairness. *Journal of Experimental Child Psychology*, *105*, 130-137.

Ting, F., He, Z., & Baillargeon, R. (2019). Where is the wide toy hidden? Five-month-old infants track what others can infer based on what they see. Manuscript in preparation.

Tomasello, M. (1999). Having intentions, understanding intentions, and understanding communicative intentions. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 63-75). Mahwah, NJ: Erlbaum.

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12-and 18-month-olds know what is new for other persons. *Developmental Psychology*, *39*, 906.

Vouloumanos, A., Martin, A., & Onishi, K. H. (2014). Do 6-month-olds understand that speech can communicate? *Developmental Science, 17*, 872-879.

Wang, S., Baillargeon, R., Brueckner, L. (2004). Young infants' reasoning about hidden objects: Evidence from violation-of-expectation tasks with test trials only. *Cognition*, *93*, 167-98.

Wang, L., & Leslie, A. M. (2016). Is implicit theory of mind the 'Real Deal'? The own-belief/true-belief default in adults and young preschoolers. *Mind & Language*, *31*, 147-176.

Wang, Y., & Su, Y. (2009). False belief understanding: Children catch it from classmates of different ages. *International Journal of Behavioral Development, 33*, 331–336.

Welder, A. N., & Graham, S. A. (2001). The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development*, *72*, 1653-1673.

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., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development*, *82*, 780-792.

Westra, E., & Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, *158*, 165-176.

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

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

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development*, *22*, 145-160.

Woodward, A. L. (2005). The infant origins of intentional understanding. *Advances in Child Development and Behavior*, *33*, 229-262.

Woodward, A. L., Sommerville, J. A., & Guajardo, J. J. (2001). How infants make sense of intentional action. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds), *Intentions and Intentionality: Foundations of Social Cognition* (pp. 149-169). Cambridge, MA: MIT Press.

Yazdi, A. A., German, T. P., Defeyter, M. A., & Siegal, M. (2006). Competence and performance in belief-desire reasoning across two cultures: The truth, the whole truth, and nothing but the truth about false belief? *Cognition*, *100*, 343-368.

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

Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology*, *106*, 208-220.

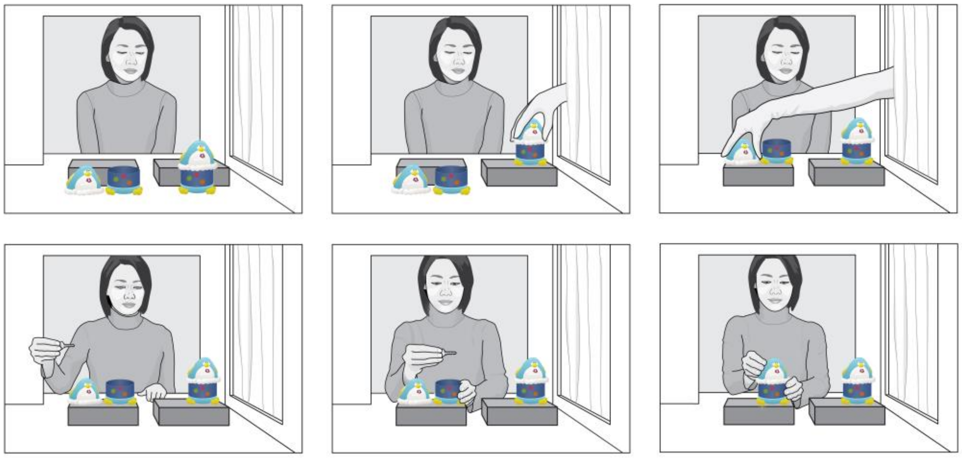
Figure Captions

Figure 1. Familiarization and test trials shown in the false-belief condition of Scott and Baillargeon (2009). The order of the two test events was counterbalanced across infants.

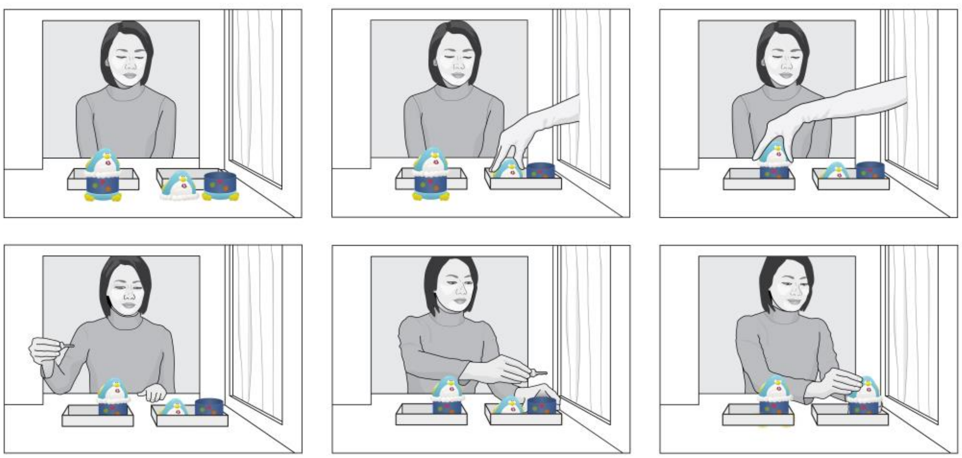
Figure 2. Familiarization and test trials shown in the deception condition of Scott, Richman, and Baillargeon (2015).

Familiarization Trials

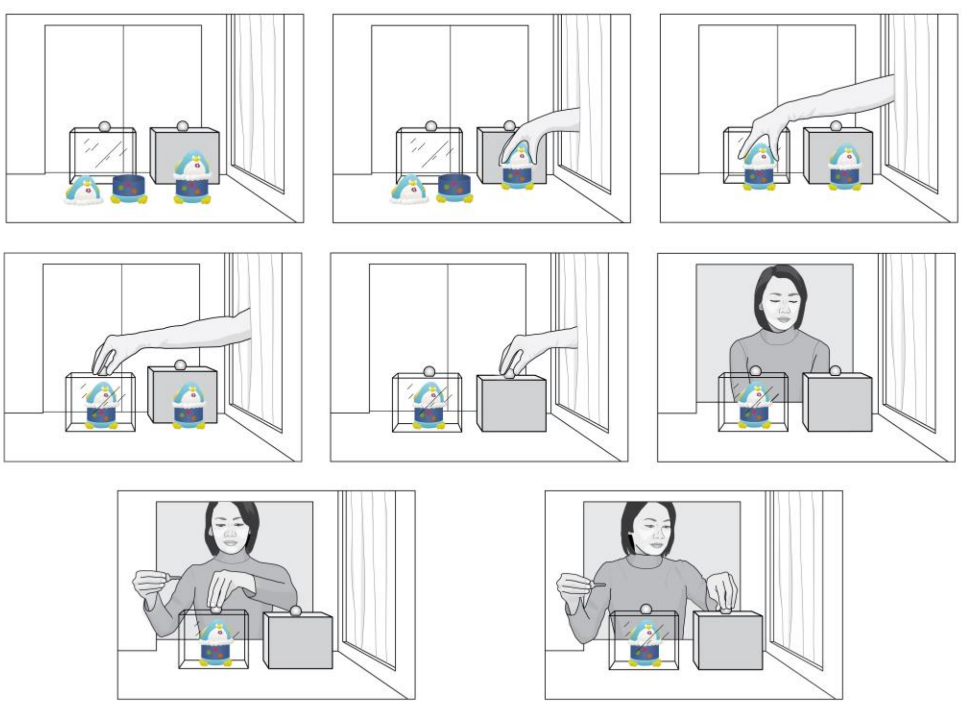
Trials 1 and 2



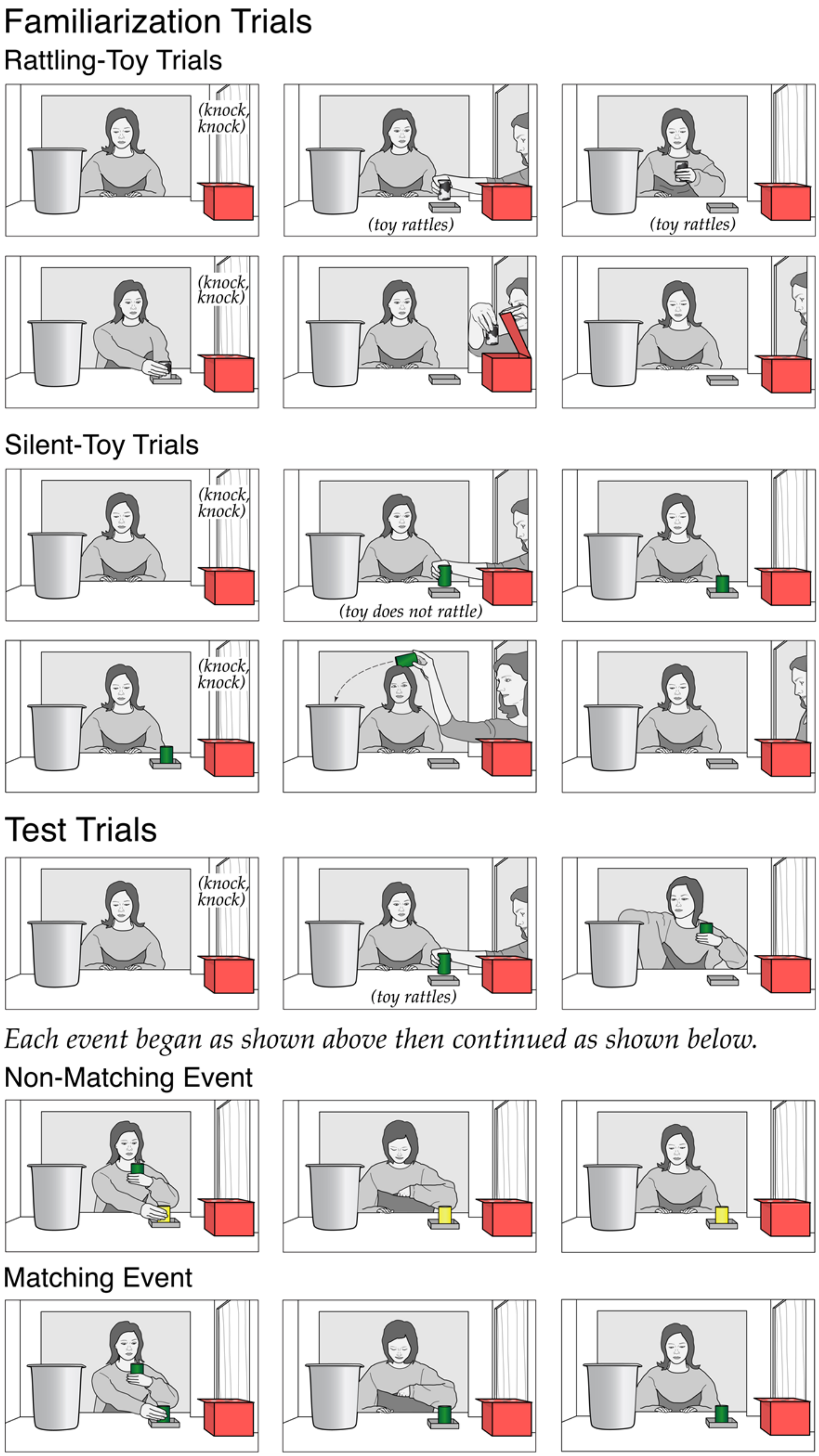
Trials 3 and 4



Test Trials



Transparent-Cover Event   Opaque-Cover Event


Familiarization Trials
Rattling-Toy Trials
(knock, knock)
(toy rattles)
(toy rattles)
(knock, knock)
Silent-Toy Trials
(knock, knock)
(toy does not rattle)
(knock, knock)
Test Trials
(knock, knock)
(toy rattles)
Each event began as shown above then continued as shown below.
Non-Matching Event
Matching Event