False-belief understanding in the first years of life

Rose M. Scott
Erin Roby
Megan Smith

In J. Kiverstein (Ed.), Routledge Handbook of the Philosophy of the Social Mind (pp. 152-171).

Please address correspondence to:
Rose Scott
University of California Merced
5200 North Lake Road
Merced, CA 95343
rscott@ucmerced.edu
As members of a social species, a large part of our everyday lives involves predicting, interpreting, and responding to the behavior of other individuals. Adults typically do this by considering others’ underlying mental states. Thus, we readily understand that Dorothy wants to return home from Oz, does not know that the Wizard is actually just a man pretending to be a wizard, and falsely believes that he can send her home. Developmental psychologists have long been interested in how this psychological reasoning ability develops. In particular, considerable research has focused on when children understand that others can be mistaken, or hold false beliefs, about the world. False-belief understanding provides evidence of the ability to distinguish between the mind and reality – to recognize that mental states are internal representations rather than direct reflections of the world. This sophisticated, and perhaps uniquely human (e.g., Kaminski, Call, & Tomasello, 2008; Marticorena, Ruiz, Mukerji, Goddu, & Santos, 2011), ability has been argued to play a vital role in cooperation, communication, and learning (e.g., Baillargeon et al., 2013; Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2007; Sperber & Wilson, 1995).

When and how does false-belief understanding develop? For several decades, this question was investigated using elicited-response false-belief tasks, in which children were asked direct questions about an individual who held a false belief (for a review, see Wellman, Cross, & Watson, 2001). These tasks are described as elicited-response tasks because the experimenter explicitly asks the child to predict or explain the behavior of the mistaken individual. In one such task (Baron-Cohen, Leslie, & Frith, 1985), children hear a story enacted with props: Sally puts her marble in a box and then leaves the room. In her absence, her friend Anne moves the marble to a nearby basket. Children are then asked where Sally will look for her marble when she returns. Beginning at around age four, children typically answer that Sally will look in the box,
where she falsely believes the marble to be. In contrast, younger children indicate that Sally will look in the basket, the marble’s current location, failing to demonstrate an understanding of Sally’s false belief. This shift from below- to above-chance performance has been widely replicated using several different elicited-response false-belief tasks (e.g., Gopnik & Astington, 1988; Perner, Leekam, & Wimmer, 1987; Wellman et al., 2001) with children around the world, although above-chance performance is not attained until age 7 in some cultures (e.g., Lecce & Hughes, 2010; Liu, Wellman, Tardif, & Sabbagh, 2008; Mayer & Träuble, 2013; Naito & Koyoma, 2006; Vinden, 2002). These findings led many to conclude that false-belief understanding constituted a major milestone in children’s psychological reasoning that was not attained until at least four years of age (Carlson & Moses, 2001; de Villiers & de Villiers, 2003; Gopnik & Wellman, 1994; Perner, 1995).

This conclusion was challenged, however, by Onishi and Baillargeon’s (2005) groundbreaking discovery that young infants demonstrated false-belief understanding when tested via other means. This study launched a new wave of research that has yielded substantial evidence of false-belief understanding in the first three years of life, shed light on the nature and extent of this early understanding, and identified factors that affect young children’s performance in false-belief tasks. In this chapter, we review this past decade of research and present theoretical accounts that have been offered to explain these findings. Finally, we discuss the broader implications of these findings for children’s social and cognitive development.

**False-belief understanding in infants and toddlers**

The first evidence that infants could attribute false beliefs to agents came from the seminal work of Onishi and Baillargeon (2005), who tested 15-month-old infants in a violation-of-expectation version of the Sally/Anne task. Infants first saw a familiarization trial where a
toy watermelon sat on the floor of a puppet stage between two boxes, one yellow and one green. A female agent entered, played with the watermelon, and then placed it inside the green box. The agent paused with her hand inside the green box until infants looked away and the trial ended. In the second and third familiarization trials, the agent entered and reached into the green box, as if grasping the watermelon, and then paused. Next, infants viewed one of several belief-induction trials. For instance, in the FB-green condition, the agent was absent and the watermelon moved from the green box to the yellow box; in the TB-yellow condition, the agent watched as the watermelon moved to the yellow box. Following the belief-induction trial, infants viewed a single test trial in which the agent reached into either the yellow (yellow-box event) or the green box (green-box event) and paused. Infants in the FB-green condition looked reliably longer if they received the yellow-box event than if they received the green-box event, suggesting that they attributed to the agent a false belief that the watermelon was in the green box, expected her to reach there in order to obtain the watermelon, and looked longer when she instead reached into the yellow box. Infants in the TB-yellow condition exhibited the opposite pattern: they expected the agent to reach into the yellow box, where she knew the watermelon was located, and looked longer if she reached into the green box instead. Together with the results of several additional conditions, these findings suggested that by 15 months of age, infants expect agents to act in accord with their beliefs, regardless of whether those beliefs are true or false.

Subsequent investigations have replicated and extended Onishi and Baillargeon’s (2005) findings in a number of ways. Several studies have confirmed that infants’ performance in false-belief tasks reflects an understanding of false belief rather than a simpler capacity to reason about ignorance (e.g., Southgate, Senju, & Csibra, 2007; Wellman, 2010) by demonstrating that infants respond differently to ignorant and mistaken agents (e.g., He, Bolz, & Baillargeon, 2011;
Knudsen & Liszkowski, 2012a; Scott & Baillargeon, 2009; Scott, Baillargeon, Song, & Leslie, 2010; Scott, Richman, & Baillargeon, 2015). If an agent is merely ignorant about an object’s location – she does not know whether it is in location-A or location-B – then infants hold no specific expectation about where she should search and look equally regardless of whether she reaches to location-A or location-B (e.g., Scott & Baillargeon, 2009; Scott et al., 2010). If, however, the agent holds a false belief that the object is in location-A, infants expect her to search in location-A and look longer if she searches in location-B instead (Onishi & Baillargeon, 2005). Thus, infants reason about ignorance and false belief as distinct mental states.

Infants can attribute a variety of false beliefs to agents, including false beliefs about the presence (e.g., Kampis, Parise, Csiibra, & Kovács, 2015; Kovács, Téglás, & Endress, 2010; Southgate & Vernetti, 2014), location (e.g., Buttelmann, Carpenter, & Tomasello, 2009; Song, Onishi, Baillargeon, & Fisher, 2008; Southgate et al., 2007; Surian, Caldi, & Sperber, 2007; Träuble, Marinović, & Pauen, 2010), identity (Buttelmann, Suhrke, & Buttelmann, 2015; Scott & Baillargeon, 2009; Scott et al., 2015; Song & Baillargeon, 2008), contents (Buttelmann, Over, Carpenter, & Tomasello, 2014), and non-obvious properties (Scott et al., 2010) of objects. For instance, Scott and Baillargeon (2009) examined 18-month-olds’ ability to reason about a false belief about an object’s identity. The experiment involved two toy penguins that were identical except that one could come apart (2-piece penguin) and one could not (1-piece penguin). In each familiarization trial, a female agent watched as an experimenter placed the 1-piece penguin and the disassembled 2-piece penguin on platforms or in shallow containers. The agent then put a key in the bottom piece of the 2-piece penguin and stacked the two pieces; the two penguins were then visually indistinguishable. In the test trials, the agent was initially absent. The experimenter assembled the 2-piece penguin, covered it with a transparent cover, and then covered the 1-piece
penguin with the opaque cover. The agent then returned with her key and reached for one of the covers. Infants looked reliably longer when the agent reached for the transparent cover, suggesting they expected her to have a false belief that the penguin under the transparent cover was the 1-piece penguin, and hence to falsely believe that the 2-piece penguin was under the opaque cover. This pattern reversed if the agent was present throughout the test trials.

In addition to understanding a range of false beliefs, infants can predict and interpret a variety of belief-based responses produced by a mistaken agent. This includes physical action responses, such as where the agent will reach (e.g., Luo, 2011; Song & Baillargeon, 2008; Senju, Southgate, Snape, Leonard, & Csibra, 2011; Träuble et al., 2010), verbal responses, such as the intended referent of a mistaken agent’s utterance (e.g., “There’s a sefo in this box”; Southgate, Chevallier, & Csibra, 2010), and the emotional responses that a mistaken agent should produce when she discovers her false belief (e.g., Moll, Kane, & McGowan, in press; Scott, 2015). In this last study (Scott, 2015), 20-month-old infants were tested in a violation-of-expectation task involving two familiarization trials and a single test trial. In the first familiarization trial, a female agent (A1) created two rattles by placing marbles inside two cups and closing them with lids. She then shook each cup in turn, demonstrating that both produced a rattling sound. In the next familiarization trial, A1 was absent; in her absence, a male agent (A2) removed the lid from one of the cups, took the marbles, and replaced the lid. A2 then repeatedly shook each of the cups in turn, demonstrating that the emptied cup no longer rattled but the other cup still did so. Next, infants received either a consistent or an inconsistent test trial. At the start of the consistent test trial, A1 was again present; A2 and the stolen marbles were absent. A1 reached for the noisy cup, shook it (it rattled), and produced a satisfied expression. She then shook the silent cup and expressed surprise. A1 continued to shake the cups in turn, producing the corresponding facial
expressions, until infants looked away. In the inconsistent trial, the pairings of cups and facial expressions were reversed: A1 was surprised by the noisy cup and satisfied with the silent cup. Infants looked significantly longer if they received the inconsistent rather than the consistent trial, suggesting they attributed to A1 the false belief that both cups rattled, expected her to be surprised when she shook the silent cup and discovered she was mistaken, and therefore looked longer if she was surprised by the noisy cup instead (this effect was eliminated if A1 knew that one of the cups had been manipulated).

Infants can demonstrate their false-belief understanding in a variety of paradigms that assess a range of infant responses. As illustrated above, infants as young as 11 months of age succeed in violation-of-expectation tasks, looking longer when agents act in ways that are inconsistent (as opposed to consistent) with their false beliefs (e.g., Luo, 2011; Onishi & Baillargeon, 2005; Song & Baillargeon, 2008; Surian et al., 2007; Träuble et al., 2010). By 18 months of age, infants also visually anticipate where a mistaken agent will search for an object (e.g., anticipatory-looking tasks; Clements & Perner, 1994; Garnham & Ruffman, 2001; Senju et al., 2011; Southgate et al., 2007; Surian & Geraci, 2012), spontaneously point to inform a mistaken agent that an object has been moved in her absence (e.g., anticipatory-pointing tasks; Knudsen & Liszkowski, 2012a, 2012b), and use an agent’s false belief to guide prompted responses such as retrieving an object for the agent (e.g., elicited-intervention tasks; Buttelmann et al., 2009; Buttelmann et al., 2014; Buttelmann et al., 2015; Rhodes & Brandone, 2014; Southgate et al., 2010).

For instance, in the elicited-intervention task devised by Buttelmann et al. (2009), an experimenter showed 18-month-old infants how to lock and unlock two boxes, leaving them unlocked. A male agent then entered and showed the infants a toy. He placed the toy into one of
the boxes and left the room. While he was gone, the experimenter moved the toy to the other box and locked both boxes. The agent then returned and attempted to open the (now empty) box where he had previously hidden the toy. When he could not open it, he sat down between the two boxes looking disappointed. If infants did not spontaneously intervene, the experimenter encouraged the infant by saying, “Go on, help him.” Most infants approached the box that contained the agent’s toy, suggesting that they interpreted the agent’s actions on the empty box as an attempt to retrieve his toy, which he falsely believed was in that location, and therefore they opened the box that held the toy in an attempt to help the agent achieve his goal. This pattern reversed if the agent saw the experimenter move the toy to the other location. In this case, infants approached and opened the empty box, suggesting that they assumed the agent’s goal was to open the box he acted on rather than to retrieve the toy. These results have now been extended to helping scenarios in which the agent holds a false belief about the contents (e.g., Buttelmann et al., 2014) or identity (e.g., Buttelmann et al., 2015) of an object.

Recent studies using neurological measures have provided converging evidence that young infants represent agents’ beliefs (e.g., Kampis et al., 2015; Southgate & Vernetti, 2014). Infants and adults activate regions of the motor cortex when generating predictions about an agent’s actions (Southgate & Begus, 2013; Stadler et al., 2012). Southgate and Vernetti (2014) thus used changes in motor cortex activation, as measured by EEG, to assess 6-month-olds’ ability to predict the actions of a mistaken agent. Infants viewed two types of test trials in which an agent sat behind a closed box. In false-belief-present test trials, the lid of the box opened, then a ball rolled on screen and jumped into the box. The lid closed and a curtain came down in front of the agent. The ball then jumped out of the box and rolled off screen. The curtain then went up to reveal the agent looking down at the closed box. The agent remained stationary for 1500ms
before reaching for the lid of the box. In false-belief-absent trials, the ball initially jumped out of the box and rolled off screen. The curtain lowered in front of the agent and the ball returned and jumped into the box. The curtain then went up and the agent remained stationary. Analysis of motor cortex activation during the 1500ms period after the curtain was raised revealed a significant increase in motor activation during the false-belief-present trials but not during the false-belief-absent trials. This selective increase in motor activation indicates that infants expected the agent to reach for the box when she falsely believed it contained a ball but not when she falsely believed the box was empty.

Finally, positive evidence of false-belief understanding has been obtained with infants and toddlers in numerous Western countries (e.g., Buttelmann et al., 2009; Kovács et al., 2010; Meristo et al., 2012; Southgate et al., 2007; Surian et al., 2007). Recently, Barrett and colleagues extended these results to include a Salar community in western China, a Shuar/Colono community in Ecuador, and a Yasawan community in Fiji (Barrett et al., 2013). These three communities differ from one another substantially in location, language, and cultural practices (see Barrett et al., 2013, supplementary online material), while sharing a number of features that distinguish them from the countries in which most early false-belief research has been conducted: they are small, rural, non-industrialized, less wealthy communities with low levels of formal education (Henrich, Heine, & Norenzayan, 2010). Barrett et al. (2013) administered three false-belief tasks that had been previously used with children in the United States (e.g., He, Bolz, & Baillargeon, 2012; Scott et al., 2010; Scott, He, Baillargeon, & Cummins, 2012) and obtained positive results in all tasks at all three sites. No differences were found across sites and in each

---

1 If the motor system were used to infer mental states, as suggested by some simulation accounts (e.g., Gallese & Goldman, 1998), then increased motor activation should have occurred in both conditions: in both cases, infants needed to infer the agent’s mental states. The selective pattern of activation observed thus suggests that motor cortex is involved in predicting what agents will do given their mental states rather than inferring those states per se (see also Jacob & Jeannerod, 2005).
case children performed quite similarly to children tested in the United States. This evidence suggests that the capacity to attribute false beliefs to others emerges universally early in development.

**Theoretical accounts of false-belief understanding**

The studies reviewed above suggest that infants can reason about a rich set of belief-based behaviors across a range of belief-inducing situations, that infants can demonstrate their understanding in a variety ways, and that this understanding emerges in infancy across cultures. Yet the results of elicited-response false-belief tasks suggest that the capacity to represent beliefs does not emerge until at least age four and that the acquisition of this ability exhibits substantial cross-cultural variability. How can we reconcile these two conflicting sets of findings? Broadly speaking, two types of accounts have been proposed to explain this discrepancy.

**Conceptual-shift accounts**

Many researchers argue that elicited-response tasks measure a different understanding than do the tasks used with infants (Apperly & Butterfill, 2009; De Bruin & Newen, 2012; de Villiers & de Villiers, 2012; Devine & Hughes, 2014; Gopnik & Wellman, 2012; Heyes, 2014; Low, 2010; Perner, 2010; Perner & Roessler, 2012; Rakoczy, Bergfeld, Schwarz, & Fizke, 2015; Ruffman, 2014; San Juan & Astington, 2012). We refer to these accounts as conceptual-shift accounts because they share the common assumption that false-belief understanding emerges after four years of age as the result of a qualitative change in children’s understanding of the mind. These accounts maintain that young children's failure on elicited-response tasks results from an inability to understand beliefs and, correspondingly, that passing elicited-response tasks demonstrates that they have attained this understanding. Although conceptual-shift accounts disagree on the precise nature of this transition, they generally agree that false-belief
understanding is a learned, culturally-constructed skill (e.g., Heyes & Frith, 2014; Ruffman, 2014; Wellman, 2014).

Proponents of the conceptual-shift view argue that the tasks used with infants do not measure an understanding of belief and instead attribute infants’ successful performance to a variety of alternative factors. Some assume that these tasks measure entirely non-mentalistic reasoning (e.g., Heyes, 2014; Perner, 2010; Ruffman, 2014), while others assume that infants are capable of mentalistic reasoning of a more minimal, rudimentary sort (e.g., Apperly & Butterfill, 2009; Low, 2010).

Non-mentalistic accounts. Two non-mentalistic accounts have been offered for infants’ performance in false-belief tasks. According to the low-level process account (Heyes, 2014), the tasks used with infants do not measure a capacity to represent mental states but rather reflect the operation of domain-general processes such as perception, attention, and memory. This account assumes that infants’ responses in false-belief tasks are driven primarily by low-level perceptual novelty: infants look longer at events that, relative to other recent events, have novel spatiotemporal relations amongst actions and objects. Heyes (2014) has argued that responses to perceptual novelty and responses based on mental states have been consistently confounded in the tasks administered to infants, giving rise to apparent false-belief understanding (for critical discussion of this claim, see Scott & Baillargeon, 2014). For instance, the infants in the FB-green condition of Onishi and Baillargeon (2005) might have looked longer when the agent reached to the yellow box because this was perceptually novel (she had not reached there before) rather than because this was inconsistent with her false belief.

The behavioral-rule account argues that the tasks used with infants assess expectations about behavior rather than an understanding of mental states (Perner, 2010; Ruffman, 2014). In
everyday life, children gather information, in the form of statistical regularities or behavioral rules, about how agents typically behave in particular situations. When children observe an agent in one of these situations in a laboratory task, they retrieve the appropriate behavioral rule and use it to interpret or predict the agent’s actions. For instance, it has been argued that infants expect agents to look for objects where they last saw them (Perner & Ruffman, 2005) and this expectation gives rise to a variety of responses such as anticipatory looks towards the location where an agent last saw an object (e.g., Southgate et al., 2007), and looking longer when an agent searches for an object in a location other than where she last saw it (e.g., Onishi & Baillargeon, 2005).

Detailed critiques of these accounts have been offered elsewhere, and a full consideration of those arguments is beyond the scope of this chapter (e.g., Apperly & Butterfill, 2009; Carruthers, 2013; Christensen & Michael, 2016; Jacob, 2013; Scott, 2014; Scott & Baillargeon, 2014). Instead, we merely point out that although these accounts might be able to explain the results of isolated experimental conditions, neither can account for the wealth of evidence for false-belief understanding (and psychological reasoning more generally) in infancy. For instance, the low-level novelty account cannot explain the results of paradigms that assess responses other than looking time, such as the helping task described earlier (e.g., Buttelmann et al., 2009, 2014, 2015; see also Knudsen & Liszkowski, 2012a; Moll et al., in press; Southgate et al., 2010), and both accounts have difficulty explaining cases where infants display differential responses to identical test events (e.g., Scott et al., 2015; Senju et al., 2011; for discussion see Scott, 2014; Scott & Baillargeon, 2014). Such findings cast doubt on both of these non-mentalistic accounts in their present forms.
**Minimalist accounts.** Perhaps the most influential minimalist account is the two-systems view advocated by Apperly and colleagues (e.g., Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; Low, Drummond, Walmsley, & Wang, 2014; Low & Watts, 2013). According to this account, humans possess two distinct systems for psychological reasoning: a late-developing system that emerges around age four, and an early-developing system that is present in infancy. The two systems exist in parallel in adulthood, each guiding different sets of responses. The late-developing system is capable of representing beliefs as such and is required for explicit judgments about others’ behavior. Its emergence thus enables successful responding on elicited-response false-belief tasks. This system is highly flexible, allowing children and adults to represent any belief that they themselves can entertain. But this flexibility comes at the cost of efficiency: the late-developing system is slow, effortful, and dependent on language and executive function resources.

The early-developing system is incapable of representing beliefs as such. Instead, it tracks simpler belief-like “registrations”: an agent who encounters an object registers its location and properties. Registrations can be used to interpret and predict an agent’s actions, and this allows infants to succeed in many false-belief tasks. For example, if an agent encounters an object in location-A and then it is moved to location-B in her absence, the early-developing system can predict that the agent will search for the object in location-A because this is where she last registered it (e.g., Onishi & Baillargeon, 2005; Southgate et al., 2007). Because this system represents simpler states, its operation is fast, automatic, and independent of language and executive function. This system enables psychological reasoning in infants, who have limited language and executive function resources, and guides rapid, automatic responses (e.g., anticipatory looking) in older children and adults.
This account predicts that the performance of the early-developing system should exhibit a number of “signature limits.” First, registrations can only capture physical relations between an agent and an object. As a result, the early-developing system should fail in situations that require reasoning about how an agent construes an object, such as those involving false beliefs about identity (e.g., Low & Watts, 2013). Second, because the early-developing system is encapsulated from other cognitive processes, it should be incapable of handling situations that place considerable demands on executive function recourses, such as those involving complex, interlocking sets of mental states that interact causally (e.g., Butterfill & Apperly, 2013; Low et al., 2014).

Contrary to these predictions, a growing body of evidence suggests that infants can reason about situations involving false beliefs about identity2 (e.g., Buttelmann et al., 2015; Scott & Baillargeon, 2009; Scott et al., 2015; Song & Baillargeon, 2008) and complex causal interactions amongst mental states (e.g., Choi & Luo, 2015; Moll et al., in press; Scott, 2015; Scott & Baillargeon, 2009; Scott et al., 2015). For instance, Scott et al. (2015) examined whether 17-month-olds could reason about the actions of a deceptive agent who sought to lure another agent into holding a false belief about the identity of an object. In several experiments, the thief attempted to steal a desirable rattling toy during its owner’s absence by substituting a less desirable silent toy. Infants realized the thief could only succeed if the silent toy was visually identical to the rattling toy and the owner did not routinely shake her toy when she returned. When these conditions were met, infants expected the owner to hold a false belief that the silent toy was the rattling toy.

---

2 These findings are at odds with several recent studies reporting that adults fail certain anticipatory-looking tasks involving false beliefs about identity (e.g., Low & Watts, 2013; Low et al., 2014; Wang, Hadi, & Low, in press). However, these negative findings likely reflect task demands rather than an inability to process false beliefs about identity (see Carruthers, in press-a, in press-b; Scott et al., 2015).
These findings suggest that by 17 months, infants’ psychological reasoning does not exhibit either of the signature limits thought to characterize the early-developing system. Although these findings cannot rule out the possibility that humans possess two systems for psychological reasoning, they cast strong doubt on the two-systems account in its current form (for additional arguments against this account, see Carruthers, in press-a, in press-b; Christensen & Michael, 2016; Helming, Strickland, & Jacob, in press; Michael & Christensen, in press; Thompson, 2014).

**Mentalistic accounts**

In contrast to conceptual-shift accounts, mentalistic accounts assume that infants’ successful performance in false-belief tasks does demonstrate an understanding of belief (e.g., Carruthers, 2013; Gergely, 2010; Helming, Strickland, & Jacob, 2014; Leslie, 2005; Luo & Baillargeon, 2010; Mitchell, Currie, & Ziegler, 2009; Southgate et al., 2007). Although several mentalistic accounts have been proposed (e.g., Baron-Cohen, 1995; Carruthers, 2013; Gergely & Csibra, 2003; Johnson, 2005; Leslie, 1994; Premack & Premack, 1995; Spelke & Kinzler, 2007), we focus here on the account offered by Baillargeon, Scott, and colleagues (e.g., Baillargeon, Scott, & He, 2010; Baillargeon et al., 2015; Scott & Baillargeon, 2009), which builds on a large body of evidence that infants successfully reason about a variety of mental states in the first years of life (for a review, see Baillargeon et al., 2015).

According to this mentalistic account, infants possess an evolved, domain-specific psychological-reasoning system that contains a skeletal causal framework for interpreting the behavior of agents. This system, which operates largely outside of conscious awareness, is triggered whenever infants attempt to interpret the behavior of an entity that they construe as an agent (for discussion of how infants categorize entities as agents, see Baillargeon, Scott, & Bian,
in press). The system then enables infants to infer some of the likely mental states that underlie the agent’s actions. Specifically, the psychological-reasoning system allows infants to attribute at least three different types of mental states to agents: *motivational states*, which capture the agent’s motivation in the scene (e.g., goals, preferences), *epistemic states*, which reflect the knowledge that an agent possesses or lacks about a scene, and *counterfactual states*, which include any false or pretend beliefs that the agent holds about the scene. In addition to these mental states, the psychological-reasoning system is thought to include constraints and principles that allow infants to predict how any agent ought to behave, given his or her mental states. One such principle is the *principle of rationality*, which states that all other things being equal, agents should expend as little effort as possible in order to achieve their goals (e.g., Gergely & Csibra, 2003; Gergely, Nádasdy, Csibra, & Bíró, 1995; Scott & Baillargeon, 2013; Southgate, Johnson, & Csibra, 2008).

The psychological-reasoning system is presumably quite skeletal at first, much like other domain-specific systems that have been proposed for physical and biological reasoning (e.g., Baillargeon et al., 2012; Gelman, 1990). As a result, infants’ early expectations about the behavior of agents are likely highly abstract and lacking in mechanistic detail (e.g., Keil, 1995, 2006). Infants may also face situations where they lack sufficient knowledge to infer an agent’s mental states. For instance, an infant who observes a parent typing on a laptop might very well view this action as deliberate and intentional while being unable to discern what the goal of this action might be. Consistent with this possibility, infants’ ability to infer the goal of various object-directed actions improves considerably over the first year of life. 6-month-olds who observe an agent repeatedly reach for object-A instead of object-B readily infer that the agent has the goal of obtaining object-A and expect her to reach for it in the future (e.g., Luo &
Baillargeon, 2005; Spaepen & Spelke, 2007; Woodward 1998). However, they have difficulty inferring the goal of other actions, such as pointing or looking at object-A (Kim & Song, 2008; Woodward, 2003; Woodward & Guajardo, 2002). Younger, 3-month-old infants have difficulty inferring the goal of a repeated grasping action and require additional support, such as prior experience grasping objects themselves, in order to do so (e.g., Sommerville, Woodward, & Needham, 2005).

The causal framework of the psychological-reasoning system thus provides infants with a starting point for learning about the behaviors of agents. As infants observe and interact with agents, this enriches their psychological-reasoning system, enabling them to reason about agents more effectively in the future (see also Carruthers, 2013; Christensen & Michael, 2016). Much as everyday environments provide sufficient exposure to patterned light to support normal development of the visual system, it seems likely that infants routinely obtain sufficient exposure to agents to support development of the psychological-reasoning system. Recent evidence that infants in a wide range of cultures successfully reason about agents’ motivational, epistemic, and counterfactual states supports this prediction (e.g., Barrett et al., 2013; Callaghan et al., 2011).

**False-belief failures: competence vs. performance.** If infants can represent false beliefs, as argued by mentalistic accounts, then why do children fail elicited-response false-belief tasks until at least age four? We argue that this pattern results from a distinction between competence and performance: as many researchers have pointed out, the capacity to represent false beliefs does not guarantee successful performance in false-belief tasks (e.g., Baillargeon et al., 2010; Bloom & German, 2000; Carruthers, 2013, in press-b; Chandler, Fritz, & Hala, 1989; Helming et al., 2014, in press; Leslie, 1994; Lewis & Osborne, 1990; Mitchell & Lacohée, 1991; Roth & Leslie, 1998; Scholl & Leslie, 2001; Siegal & Beattie, 1991; Yazdi, German, Defeyter,
& Siegal, 2006). Indeed, there are likely many factors that mediate between the capacity to represent false beliefs and successful performance in any given situation. Here we discuss a few such factors suggested by recent research.

**Attention/Motivation.** The psychological-reasoning system is triggered only when one attends to the behavior of an agent. Given that children readily orient to social stimuli from early in infancy (e.g., Farroni, Csibra, Simion, & Johnson, 2002; Farroni et al., 2005; Johnson, Dziurawiec, Ellis, & Morton, 1991), it is often assumed that children will naturally attend to agents in experimental paradigms. However, this may not always be the case. For instance, children diagnosed with autism spectrum disorder (ASD) orient less readily to social stimuli (e.g., Dawson, Meltzoff, Osterling, Rinaldi, & Brown, 1998). When compared to typically developing children, children diagnosed with ASD spend relatively more time attending to objects as opposed to people (e.g., Klin, Johnes, Schultz, Volkmar, & Cohen, 2002; Osterling, Dawson, & Munson, 2002; Swettenham et al., 1998). This atypical attention to social stimuli likely contributes to the well-established difficulties that children with ASD exhibit in tests of psychological reasoning (e.g., Klin et al., 2002), and may also compromise the development of the psychological-reasoning system. Even amongst typically developing children, events that disrupt children’s attention to the agent interfere with their performance on false-belief tasks (e.g., Rubio-Fernández & Geurts, 2013). Moreover, recent work with adults suggests that the inclination to attend to others’ mental states may be more predictive of everyday social functioning than the ability to correctly represent and infer mental states (e.g., Brandone, Werner, & Stout, 2015). Thus, the capacity to represent mental states is insufficient for
successful false-belief reasoning (or for navigating everyday social interactions): one must also be inclined to attend to and reason about other individuals.\(^3\)

**Processing demands.** Even if children attend to an agent and infer the contents of that agent’s false belief, they might still fail a false-belief task due to difficulties expressing this understanding. The processing-load account put forth by Baillargeon, Scott, and colleagues argues that children’s performance in any given false-belief task depends both on the processing demands imposed by that task and their ability to cope with those demands (e.g., Baillargeon et al., 2010; Baillargeon et al., 2015). If the processing demands of a task exceed children’s processing abilities, then they will fail the task despite their ability to represent false beliefs.

According to this account, one reason young children fail elicited-response false-belief tasks is because these tasks generally impose greater processing demands than do the tasks that are used with young infants. In particular, when children are asked the test question in standard elicited-response tasks (e.g., “Where will Sally look for her marble?”), a response-selection process is activated: children must interpret the test question, choose to answer it, and generate an appropriate response (Scott & Baillargeon, 2009; see also Mueller, Brass, Waszak, & Prinz, 2007; Saxe, Schulz, & Jiang, 2006). In standard elicited-response tasks, executing this response-selection process triggers a prepotent tendency to answer the question based on the marble’s actual location. At present, the source of this bias remains unclear: it could arise because (1) children’s own beliefs are naturally more salient, (2) explicitly mentioning the marble draws children’s attention to its actual location, (3) children misinterpret the test question as an indirect request for the marble’s actual location, (4) children misinterpret the test question as asking where Sally ought to look for her marble (or where she will eventually have to look in order to

\(^3\) Of course, attending to an agent’s behavior does not guarantee that one will successfully infer that agent’s beliefs. One might have difficulty inferring beliefs due to a lack of situational knowledge (e.g., Christensen & Michael, 2016) or processing demands (see Carruthers, in press-b).
find the marble), (5) or because children mistakenly adopt the experimenter’s perspective rather than Sally’s perspective (e.g., Baillargeon et al., 2015; Carruthers, 2013; Goldman, 2012; Hansen, 2010; Helming et al., 2014, in press; Lewis, Hacquard, & Lidz, 2012; Mitchell et al., 2009; Leslie & Polizzi, 1998; Rubio-Fernández & Geurts, 2013; Siegal & Beattie, 1991). Regardless of its source, children must inhibit this prepotent response in order to answer the test question correctly based on Sally’s false belief; this is challenging for younger children with immature inhibitory skills (e.g., Carlson & Moses, 2001). Finally, simultaneously holding in mind the agent’s false belief while planning and executing a response imposes substantial working memory demands (e.g., Freeman & Lacohée, 1995; Setoh, Scott, & Baillargeon, 2011).

The processing-load account predicts that if processing demands were sufficiently reduced, younger children might succeed in an elicited-response task. Several recent findings provide support for this prediction (e.g., Rubio-Fernández & Geurts, 2013, in press; Scott & Setoh, 2012; Setoh, et al., 2011). For instance, Setoh et al. (2011) devised a modified elicited-response task that was designed to reduce the inhibitory and working memory demands associated with response selection. 2.5-year-old children heard a false-belief story accompanied by a picture book. In each of six story trials, an experimenter turned a page of the book to reveal a new picture and recited a line of the story. The story introduced Emma who found an apple in one of two containers (e.g., a box), moved it to the other container (e.g., a bowl), 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 saw pictures of the two containers and were asked where Emma would look for her apple. Because the apple was removed to an unknown location, the prepotent response evoked by the test question should be weaker and easier for the children to inhibit. To further reduce the response-selection demands of the test trial, two practice
trials were interspersed amongst the story trials. In one, children saw an apple and a banana and were asked “Where is Emma’s apple?”; in the other, they saw a ball and a frisbee and were asked “Where is Emma’s ball?” In each case, children were required to point to the matching picture. These trials thus gave children practice interpreting a “where” question and producing an appropriate response by pointing to one of two pictures. Children performed reliably above chance in the test trial, pointing to the container Emma falsely believed held her apple. Additional experiments replicated these results and revealed that children failed if they received only one practice trial (Setoh et al., 2011) or if the practice questions (“Which one is Emma’s apple?”) differed in linguistic form from the test question (“Where will Emma look for her apple?”; Scott & Setoh, 2012). These results demonstrate that young children are easily overwhelmed by simultaneously representing an agent’s false belief and answering a question about this belief and that when these demands are reduced, they can succeed in elicited-response tasks at younger ages.

In contrast to elicited-response tasks, the tasks used with infants and toddlers have been designed to impose fewer processing demands on children, allowing them to express their false-belief understanding at younger ages. In particular, tasks that assess spontaneous responses such as anticipatory-looking (e.g., Southgate et al., 2007), violation-of-expectation (e.g., Onishi & Baillargeon, 2005), or preferential-looking (e.g., Scott et al., 2012) do not involve response-selection demands: because children are not asked direct questions, no response-selection process is activated. However, this does not imply that such tasks do not involve any processing demands, nor that children’s success on such tasks is guaranteed. According to the processing-load account, children’s performance in any false-belief task is jointly determined by the processing demands imposed by that task and children’s ability to cope with those demands. Just as decreasing processing demands facilitates young children’s performance on elicited-response tasks, increasing
processing demands should *impede* young children’s performance on spontaneous-response tasks. Children’s performance on high-demand spontaneous-response tasks should be correlated with their processing skills, just as performance on elicited-response tasks is correlated with skills such as inhibitory control (e.g., Carlson & Moses, 2001) and linguistic ability (e.g., Milligan, Astington, & Dack, 2007).

Scott and Roby (2015) recently tested these predictions using a novel high-demand preferential-looking task. Preferential-looking tasks take advantage of the well-established tendency for children and adults to look spontaneously at images that match the sentences they hear (e.g., Scott et al., 2012; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this task, 3-year-old children heard a change-of-location false-belief story accompanied by a picture book. The story introduced Mia, who wanted to give her grandmother a cookie for her birthday. She placed the cookie in a box, then left. In her absence, her brother moved the cookie to a nearby bag. Next, children saw Mia running into the room in a coat and heard, ‘‘Hurry hurry,’ says Mia’s mom! ‘We’re leaving for Grandma’s!’ Mia puts on her coat and quickly runs in to get Grandma’s cookie.’’ This story line was ambiguous and open to two interpretations: (1) Mia runs in and hastily grabs the container that she believes contains the cookie (the box; false-belief interpretation) and (2) Mia runs in, locates the cookie, and takes the container holding it (the bag; reality interpretation). Although both interpretations were plausible, adults find the false-belief interpretation more appropriate.

In the subsequent test trial, children saw two pictures of the back of an unknown individual in a hooded coat. In one image, the individual carried the box (original-container picture) and in the other, the individual carried the bag (current-container picture). While viewing these images, children heard, ‘‘There’s Mia walking to Grandma’s. She’s carrying Grandma’s present’’ and their
looking time to each image was measured. Because children tend to look at images that match the sentences they hear, they should look longer at the individual that they thought was Mia carrying Grandma’s present. However, children never saw which container Mia selected: in order to decide which individual was Mia, they had to infer which container she must have taken based on the story. Thus, children who arrived at the more appropriate false-belief interpretation should look longer at the original-container picture, whereas those who arrived at the less appropriate reality interpretation should instead look longer at the current-container picture. Results indicated children’s performance was strongly correlated with their verbal ability: children with more advanced verbal abilities looked longer at the original-container picture, whereas those with low verbal abilities looked longer at the current-container picture. These results thus confirm that even in a spontaneous-response task, children’s performance depends on processing demands and processing skills (see also Schneider, Lam, Bayliss, & Dux, 2012; Yott & Poulin-Dubois, 2012).

**Universal origins, culturally specific endpoints.** We have argued that the capacity to represent beliefs develops universally in the first years of life as a part of the psychological-reasoning system, but that children’s performance in false-belief tasks is constrained by factors such as attention, motivation, and processing demands. This raises several possible avenues by which environmental factors could lead to individual and cultural differences in the expression of a universal capacity for representing beliefs. For instance, there is a well-established link between social input and children’s performance on false-belief tasks (e.g., Adrián, Clemente, Villanueva & Rieffe, 2005; Ensor, Devine, Marks & Hughes, 2014; Mayer & Träuble, 2013; Mayer & Träuble, 2013; Yott & Poulin-Dubois, 2012).

---

4 When presented with a single scene, children tend to respond with increased attention if that scene violates their expectations (i.e. violation-of-expectation tasks). However, several decades of research has shown that when children view multiple images accompanied by a word or sentence, they tend to look at the image that matches the language they hear (e.g., Golinkoff, Ma, Song, & Hirsh-Pasek, 2013). Thus, the prediction here was that children would look at the image that was consistent with their interpretation of the story, rather than the image that was ‘unexpected’ based on the story (see also Scott et al., 2012).
Meins et al., 2003; Ruffman, Slade, & Crowe, 2002; Taumoepeau & Reese, 2013). Children whose mothers use more mental-state terms in conversation (especially cognitive terms such as think and know) perform better on elicited-response false-belief tasks (e.g., Adrián et al., 2005; Ensor & Hughes, 2008; Ruffman et al., 2002; Symons, Peterson, Slaughter, Roche & Doyle, 2005). In cultures where the discussion of mental states is less socially appropriate, children pass elicited-response tasks at later ages (Mayer & Träuble, 2013) and they also perform more poorly on other tests of social understanding (Taumoepeau, Reese, & Gupta, 2012). Recent evidence suggests that the relationship between social input and false-belief performance also holds for spontaneous-response false-belief tasks. 2.5-year-olds who hear more mental-state terms more readily anticipate where a mistaken agent will search for a goal object (Roby & Scott, 2015). Conversely, deaf infants of hearing parents, who are exposed to significantly less mental-state language than hearing infants of hearing parents (Morgan et al., 2014), fail anticipatory-looking false-belief tasks at 17 months of age (Meristo et al., 2012).

These studies demonstrate that from early in infancy, social input is related to children’s performance in both spontaneous- and elicited-response false-belief tasks. One possibility is that children who frequently engage in conversations about mental states more readily attend to the mental states of others. Alternatively, frequent discussion of mental states could provide children with practice inferring the content of others’ mental states, which would then facilitate children’s false-belief inferences in both spontaneous- and elicited-response tasks. While further research is needed to disentangle these possibilities, these findings demonstrate one way in which environmental/cultural differences interact with a universal capacity for belief representation to yield individual differences in false-belief performance.5

---

5 In addition to affecting the expression of children’s intuitive psychological reasoning, culture also certainly impacts the creation of an explicit ‘folk theory’ of other’s minds (for discussion, see Lillard, 1998).
Implications for children’s social and cognitive development

As discussed at the outset of this chapter, understanding the origins of false-belief reasoning has implications beyond theoretical debates about the cognitive mechanisms that drive psychological reasoning. The capacity to represent beliefs has been argued to play a fundamental role in both the development and evolution of many human social behaviors (e.g., Baillargeon et al., 2013; Sperber & Wilson, 1995; Hermann et al., 2007). For instance, the ability to recognize that others’ representations of the world may differ from our own may allow us to more effectively interpret the communicative behaviors of conversational partners and to tailor our own communicative behaviors to meet the needs of our addressee (e.g., Brown-Schmidt, 2009; Hanna, Tanenhaus, & Trueswell, 2003; Lockridge & Brennan, 2002; Nadig & Sedivy, 2002; Shwe & Markman, 1997; Southgate et al., 2010; Tager-Flusberg, 2000). This is essential to cultural learning, where novice learners’ assumptions and beliefs may be very different from those of their teachers. False-belief reasoning might also facilitate cooperation and living in large social groups by allowing us to distinguish between what we think and feel internally and what we convey externally to others (e.g., Baillargeon et al., 2013).

A number of studies attest to the importance of false-belief understanding for other aspects of development: children’s performance on elicited-response false-belief tasks is correlated with their social competence (Lalonde & Chandler, 1995), peer acceptance (Fink, Begeer, Peterson, Slaughter, & de Rosnay, 2015; Slaughter, Dennis, & Pritchard, 2002), and cooperation in peer interactions (Dunn & Cutting, 1999). However, the causal relationship between false-belief understanding and these developmental outcomes remains unclear and depends in part on how one interprets recent positive evidence of false-belief understanding in infancy. If successful performance on elicited-response tasks indicates the emergence of a
representational understanding of the mind, then these associations suggest that developing the capacity to represent others’ minds leads to advances in a variety of social behaviors (e.g., Wellman, 2014). If, however, the capacity to represent others’ minds emerges universally in infancy, then a different explanation of these correlations is warranted. Clearly, elicited-response tasks measure an ability that is broadly related to other aspects of development. If this is not the capacity to represent others’ minds per se, then what might it be? Perhaps these correlations reflect important individual differences in attention to, or skill at inferring the contents of, others’ mental states. If so, then performance on spontaneous-response tasks might also correlate with skills such as social competence and cooperation. Alternatively, these associations might reflect the importance of individual differences in children’s ability to express their false-belief understanding while coping with other demands. Scott and Baillargeon (2009) suggested that successful performance in elicited-response false-belief tasks might reflect maturation of neural connections between brain regions devoted to representing beliefs and those devoted to response selection. The maturation of such pathways might also support the expression of false-belief understanding in social situations, giving rise to the associations described above.

Clarifying these issues has important implications for supporting optimal social cognitive development in children. For instance, some research suggests that preschool children from lower socioeconomic backgrounds perform more poorly on elicited-response tasks than their more advantaged peers (e.g., Cutting & Dunn, 1999; Holmes, Black, & Miller, 1996; Hughes et al., 2000) and this may place them at risk for poorer outcomes, including poorer social competence. These children might benefit from interventions intended to improve their prospects. However, what skills should such an intervention target? This depends on the underlying cause of the relationship between elicited-response tasks and developmental
outcomes. Thus, additional research is needed to clarify the abilities assessed by different types of false belief tasks as well as how these abilities relate to other areas of development.
References


Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned? Perspectives on Psychological Science, 8, 316-339.


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


