The Developmental Origins of False-Belief Understanding

Rose M. Scott

Corresponding Author:

Rose M. Scott, Psychological Sciences, University of California, Merced, 5200 North Lake Rd., Merced, CA 95343

E-mail: rscott@ucmerced.edu

Abstract

Understanding that individuals can be mistaken, or hold false beliefs, about the world is an important human ability that plays a vital role in social interactions. When and how does this ability develop? Traditional investigations using elicited-response tasks suggested that false-belief understanding did not emerge until at least age 4. However, more recent studies have shown that children demonstrate false-belief understanding much earlier when tested via other means. In the present article, I summarize recent evidence that a robust, flexible understanding of false belief emerges in infancy and discuss why older children fail elicited-response tasks despite their ability to represent beliefs.

Keywords

cognitive development, false belief, infancy, psychological reasoning

Much of our everyday lives involves predicting, interpreting, and responding to the behavior of *agents*, entities that can perceive their environment and control their actions. Adults do this by considering agents' mental states, including their goals, knowledge, and beliefs. Developmental psychologists have long been interested in how children develop the ability to attribute mental states to agents. In particular, considerable research has focused on when children understand that agents can be mistaken or hold false beliefs about the world. False-belief understanding requires the ability to recognize that mental states are internal representations rather than direct reflections of reality and thus can be inaccurate. This ability may be uniquely human (e.g., Kaminski, Call, & Tomasello, 2008, but see Krupenye, Kano, Hirata, Call, & Tomasello, 2016) and has been argued to play an important role in many human social behaviors (e.g., Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2007). For instance, recognizing that others' representations of the world may differ from our own helps us communicate more effectively and facilitates cooperation (e.g., Baillargeon et al., 2013).

The development of false-belief understanding was traditionally investigated using elicited-response false-belief tasks, in which children were asked direct questions that required them to predict or explain the behavior of an agent who held a false belief (for a review, see Wellman, Cross, & Watson, 2001). In one such task (Baron-Cohen, Leslie, & Frith, 1985), children hear a story enacted with props: Sally puts a toy in a container and then leaves (Fig. 1). In her absence, Anne moves the toy to another container. Children are then asked where Sally will look for her toy when she returns. Beginning around age 4, children correctly answer that Sally will look in the original container, where she falsely believes the toy to be. In contrast, younger children incorrectly answer that Sally

will look in the toy's current location, suggesting an inability to understand Sally's false belief. This pattern of findings was widely replicated using several different elicited-response tasks, leading many to conclude that false-belief understanding did not emerge until at least age 4 (Wellman et al., 2001).

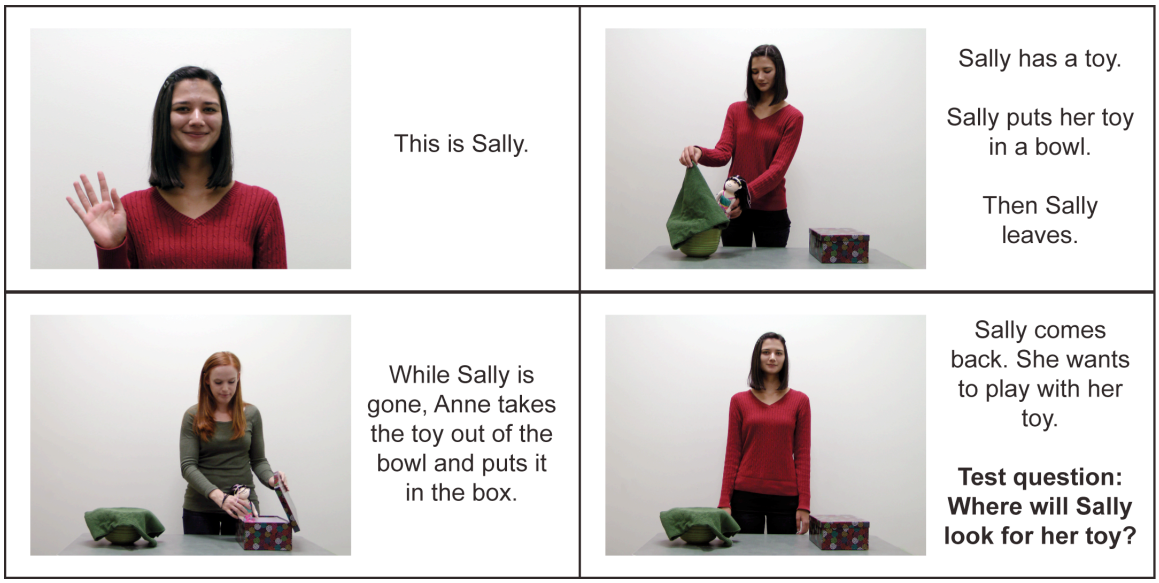

Fig. 1.

The Sally/Anne false-belief task. Children are introduced to Sally, who places a toy into one of two containers (i.e., the bowl). In her absence, Anne moves the toy to the other container (i.e., the box). Sally then returns, and children are asked where Sally will look for her toy.

This conclusion was challenged by the discovery that much younger children appear to demonstrate false-belief understanding when tested with tasks that do not involve direct questions about a mistaken agent's behavior (non-elicited-response tasks). In a seminal study, Onishi and Baillargeon (2005) tested 15-month-old infants in a violation-of-expectation task, which relies on infants' tendency to look longer at events that are inconsistent, as opposed to consistent, with their expectations. The infants saw an agent play with a toy watermelon and hide it in a green box. In the agent's absence, the

watermelon moved from the green box to a nearby yellow box. In the test trial, the agent returned and reached into either the green box (green-box event) or the yellow box (yellow-box event) and paused. The infants who saw the yellow-box event looked reliably longer than those who saw the green-box event, suggesting that they attributed to the agent a false belief that the toy was in the green box, expected her to reach into the green box in order to obtain the toy, and hence looked longer when she violated this expectation by reaching into the yellow box. Together with the results of several additional conditions, this finding suggested that by 15 months, infants realize that agents can hold and act on false beliefs.

This study launched a new wave of research into early false-belief understanding, as many researchers adopted novel approaches to studying this ability in infants and toddlers. Since 2005, over two dozen published studies have reported evidence of false-belief understanding in children between 6 months and 3 years of age (for a review, see Baillargeon et al., 2015). These subsequent studies have addressed two important, interrelated questions regarding the origins and development of false-belief understanding.

What Is the Nature of Children's Early Competence?

There has been considerable debate regarding the nature of the competence measured in non-elicited-response tasks. Advocates of *mentalistic* accounts argue that like elicited-response tasks, non-elicited-response tasks assess children's ability to attribute false beliefs to agents (Buttelmann, Carpenter, & Tomasello, 2009; Carruthers, 2013; Kovács, Téglás, & Endress, 2010; Scott, Baillargeon, Song, & Leslie, 2010; Southgate, Senju, & Csibra, 2007). On this view, positive results in non-elicited-response tasks indicate that

the capacity to attribute false beliefs to agents emerges in infancy. However, some researchers have questioned this view, arguing that responses in non-elicited-response tasks are driven by limited, rudimentary abilities, such as a tendency to look longer at perceptually novel configurations of "colors, shapes, and movements" (Heyes, 2014, p. 648), learned behavioral rules for how agents typically behave in particular situations (e.g., Ruffman, 2014), or an early-developing system for tracking the information registered by an agent (e.g., Apperly & Butterfill, 2009). According to these *late-emergence* accounts, the capacity to represent false beliefs does not emerge until age 4, as indicated by success in elicited-response tasks.

Are infants and toddlers capable of representing beliefs, or does their success in non-elicited-response tasks instead reflect more primitive capacities? One approach to addressing this question is to examine the range of non-elicited-response tasks in which young children succeed. School-aged children demonstrate false-belief understanding in a variety of elicited-response tasks involving different types of false beliefs, and they can use an agent's false belief to interpret and predict a range of belief-based responses produced by that agent (Wellman et al., 2001). This demonstrates that older children possess a robust understanding of belief that they can flexibly use in a variety of situations. If infants and toddlers are capable of attributing false beliefs to agents, then they should also demonstrate the "flexible use of belief understanding" that is seen in older children (Perner & Ruffman, 2005, p. 216). If they were instead limited to reasoning about an arbitrary subset of belief-inducing situations and belief-based responses, then this might suggest that their performance was driven by mechanisms other than an understanding of belief.

Accumulating evidence suggests that early false-belief understanding is indeed flexible and robust. Infants and toddlers demonstrate false-belief understanding in numerous different non-elicited-response paradigms that measure a range of responses, including children's looking behavior, helping responses, emotional expressions, and neurological activity (Buttelmann et al., 2009; Kovács et al., 2010; Moll, Kane, & McGowan, 2016; Southgate & Vernetti, 2014). They can use an agent's false belief to interpret that agent's actions (e.g., Luo, 2011), predict the agent's future behavior (e.g., Southgate et al., 2007; Surian & Geraci, 2012), and guide their own interactions with the agent (e.g., Buttelmann et al., 2009). For instance, in one experiment, 17-month-olds watched an agent hide two novel objects in two boxes and then leave; in her absence, an experimenter switched the objects' locations (Southgate, Chevallier, & Csibra, 2010). The agent returned, pointed to one of the boxes, and said it contained a "sefo." She then asked the infants, "Can you get the sefo for me?" Most infants approached the box that the agent had not pointed to, suggesting that they understood that she held a false belief about the boxes' contents and thus intended to label the other object as the sefo.

Like older children, infants and toddlers can represent many types of false beliefs, including false beliefs about the contents, identity, location, presence, and properties of an object (e.g., He, Bolz, & Baillargeon, 2011; Kampis, Parise, Csibra, & Kovács, 2015; Scott & Baillargeon, 2009; Scott et al., 2010). They can reason about a variety of physical actions performed by a mistaken agent, such as where the agent will search or which object the agent will select (e.g., Scott et al., 2010; Song & Baillargeon, 2008; Surian, Caldi, & Sperber, 2007). They can use an agent's false belief to interpret what

that agent says (Southgate et al., 2010) and to predict and interpret the agent's emotional reactions (e.g., Moll et al., 2016; Scott, 2017).

Moreover, young children can reason about complex causal interactions between false beliefs and other mental states (Scott & Baillargeon, 2009; Scott, Richman, & Baillargeon, 2015). For instance, Scott et al. (2015) examined 17-month-olds' ability to interpret the actions of a deceptive agent who sought to lure another agent into holding a false belief about the identity of an object. In three experiments, a thief attempted to steal a desirable rattling toy without its owner's knowledge by replacing it with a less desirable silent toy in the owner's absence. Results suggested that the infants realized that the thief could get away with her theft only if (a) the silent object was visually identical to the rattling object that the thief stole (otherwise, the owner would detect the substitution when she saw the silent toy) and (b) the owner did not routinely shake her toy when she returned (otherwise, she would detect the substitution when she shook the toy). When these conditions were met, the infants expected the owner to be deceived by the thief's actions and to falsely believe that the silent toy was the rattling toy. Thus, the infants understood the interaction between the thief's goals and the owner's false belief about the identity of the silent toy.

Finally, positive results have been obtained using non-elicited-response tasks with infants and toddlers in many Western countries, as well as in three traditional, non-Western communities: a Salar community in western China, a Shuar/Colono community in Ecuador, and a Yasawan community in Fiji (Barrett et al., 2013).

The findings reviewed here cast doubt on the notion that young children's performance in non-elicited-response tasks reflects limited, rudimentary abilities (for critiques of late-

emergence accounts, see Carruthers, 2013, 2016; Scott, 2014; Scott & Baillargeon, 2014). Instead, these findings suggest that across cultures, the capacity to attribute false beliefs to agents emerges in the first year of life. This early understanding is robust, flexible, and sophisticated, allowing young children to reason about a broad range of belief-based responses and belief-inducing situations, including complex interactions between beliefs and other mental states. Together with the large body of evidence that infants can reason about other mental states in the first year of life (e.g., Baillargeon et al., 2015), these findings support mentalistic accounts of early psychological reasoning, which propose that infants possess a psychological-reasoning system that provides a skeletal causal framework for interpreting and predicting the actions of agents in terms of their mental states, including false beliefs (e.g., Luo & Baillargeon, 2010).

Why Do Children Fail Elicited-Response False-Belief Tasks?

If the ability to attribute false beliefs to agents emerges in infancy, as the evidence reviewed in the previous section suggests, then why do children fail elicited-response false-belief tasks until age 4? Some researchers have suggested that young children fail elicited-response tasks because of the heavy processing demands that these tasks impose (e.g., Baillargeon, Scott, & He, 2010; Bloom & German, 2000; Roth & Leslie, 1998). In particular, the *processing-demands* account proposed by Baillargeon, Scott, and colleagues argues that elicited-response tasks involve at least three executive function demands (Baillargeon et al., 2010; Scott & Roby, 2015). To illustrate, consider the Sally/Anne task described earlier. When children are asked the test question, "Where will Sally look for her marble?", this initiates a *response-selection process*: Children must interpret the question, choose to answer, and select an appropriate response. Interpreting

this question often triggers an inappropriate, prepotent tendency to answer based on the marble's actual location, and children must inhibit this response in order to answer based on Sally's false belief (*response-inhibition process*). Simultaneously holding in mind the agent's false belief and executing the response-selection and response-inhibition processes imposes substantial demands on children's working memory. These combined response-selection, response-inhibition, and working memory demands overwhelm young children, causing them to fail elicited-response tasks.

The processing-demands account predicts that if processing demands are sufficiently reduced, children should succeed in elicited-response tasks at younger ages. Several recent findings support this prediction (e.g., Rubio-Fernández & Geurts, 2013; Setoh, Scott, & Baillargeon, 2011). For instance, 2.5-year-olds succeed in an elicited-response task in which (a) the location of the marble is unknown, making it easier for the children to inhibit reality-based responses, and (b) the children receive practice trials designed to reduce the demands associated with the response-selection process (e.g., Setoh et al., 2011). Note that if young children failed elicited-response tasks because they could not represent false beliefs (as argued by late-emergence accounts), then reducing processing demands would have little effect on their performance: They would continue to fail such tasks. The fact that reducing processing demands improves young children's performance thus demonstrates that they can represent false beliefs, but they are easily overwhelmed by the joint demands imposed by simultaneously doing so and answering a question about that belief.

According to the processing-demands account, non-elicited-response tasks impose fewer demands on children than do elicited-response tasks, allowing children to express their

false-belief understanding at younger ages. For instance, Scott et al. (2012) tested 2.5-year-olds in a violation-of-expectation task in which they watched an adult subject participate in a classic Sally/Anne task (Fig. 2). When the adult subject was asked where Sally would think her toy was, children expected the adult to answer correctly by pointing to the container that Sally falsely believed held the toy, and they looked reliably longer if the adult instead pointed to the toy's current location. Because the test question was addressed to the adult subject rather than the children, no response-selection or response-inhibition processes were activated, allowing children to successfully demonstrate their understanding of Sally's false belief.

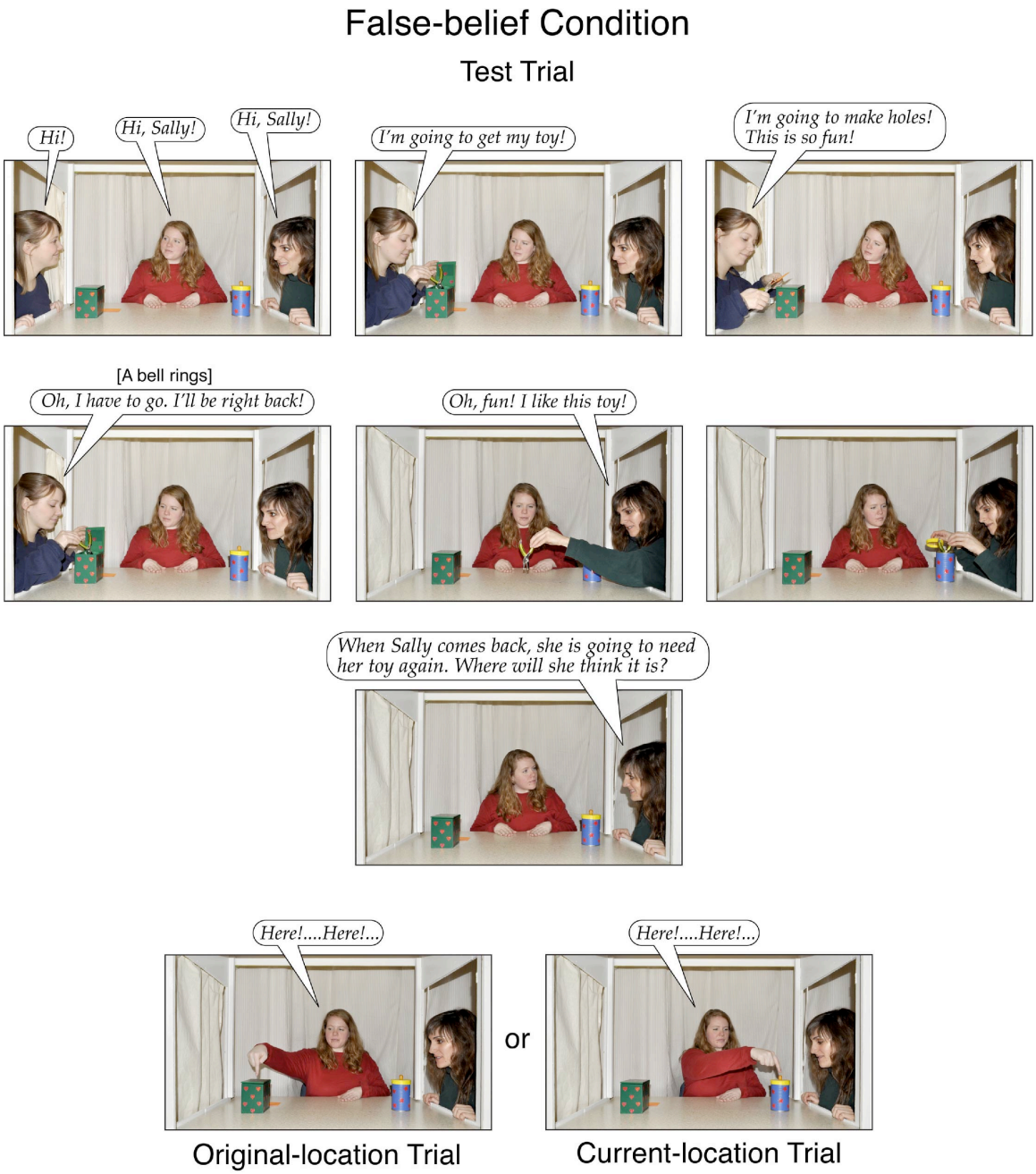

Fig. 2.

The version of the Sally/Anne task used by Scott, He, Baillargeon, and Cummins (2012). Sally (the woman on the left) places her toy into the green box and then leaves. In her absence, the experimenter (the woman on the right) moves the toy from the green box to the blue box. The experimenter then asks the adult subject (the woman in the middle) the standard Sally/Anne test question of where Sally will look for her toy when she returns.

Two-and-a-half-year-olds expected the adult subject to answer correctly that Sally's false belief would cause her to look in the green box, and they looked reliably longer if the adult instead pointed to the toy's current location.

The processing-demands account does not claim that non-elicited-response tasks do not involve *any* demands or that children should always pass such tasks. Rather, this account maintains that children's ability to demonstrate false-belief understanding in any situation should depend on their ability to cope with the demands imposed by that situation. This account thus predicts that just as decreasing processing demands improves young children's performance in elicited-response tasks, *increasing* processing demands should *impede* young children's performance in non-elicited-response tasks. Recent work by Scott and Roby (2015) supports this prediction: Three-year-olds were tested in a non-elicited-response task involving a verbal false-belief story, and the linguistic ambiguity of the story (i.e., whether story lines were open to multiple interpretations) varied across children. When the story was unambiguous, children successfully demonstrated false-belief understanding. When the story was ambiguous, however, children's performance depended on their language abilities, and only those children with the highest verbal abilities succeeded. These findings support the processing-demands account by demonstrating that regardless of whether a task involves an elicited response, children's ability to succeed depends on the processing demands of the task and their processing skills.

Recommended Reading

Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K., . . . Bian, L. (2015). (See References). A comprehensive review of psychological and socio-moral reasoning in infancy, including the development of false-belief understanding.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language, 28*, 606–637. An alternative view of the evidence discussed in this article.

Onishi, K. H., & Baillargeon, R. (2005). (See References). The first study to examine false-belief understanding in infancy.

Scott, R. M., Richman, J. C., & Baillargeon, R. (2015). (See References). A study that illustrates research on false-belief understanding in infancy and includes a detailed discussion of alternative views on this topic.

Acknowledgments

The author thanks John Bunce, Erin Roby, Megan Smith, and three anonymous reviewers for helpful feedback on the manuscript.

Declaration of Conflicting Interests

The author declared no conflicts of interest with respect to the authorship or the publication of this article.

References

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

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

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

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

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 B: Biological Sciences, 280*, 20122654.

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*, B25–B31.

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

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.

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.

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*, 1360–1366.

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

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition, 109*, 224–234.

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.

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

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science, 354,* 110-113.

Luo, Y. (2011). Do 10-month-old infants understand others' false beliefs? *Cognition, 121*, 289–298.

Luo, Y., & Baillargeon, R. (2010). Toward a mentalistic account of early psychological reasoning. *Current Directions in Psychological Science, 19*, 301–307.

Moll, H., Kane, S., & McGowan, L. (2016). Three-year-olds express suspense when an agent approaches a scene with a false belief. *Developmental Science, 19*, 208–220.

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

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science, 308*, 214–216.

Roth, D., & Leslie, A. M. (1998). Solving belief problems: Toward a task analysis. *Cognition, 66*, 1–31.

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

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

Scott, R. M. (2014). Post hoc versus predictive accounts of children's theory of mind: A reply to Ruffman. *Developmental Review, 34*, 300–304.

Scott, R. M. (2017). Surprise! 20-month-old infants understand the emotional consequences of false beliefs. *Cognition, 159*, 33–47.

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

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

Scott, R. M., 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. doi:10.1371/journal.pone.0142405

Setoh, P., Scott, R. M., & Baillargeon, R. (2011, March). *False-belief reasoning in 2.5-year-olds: Evidence from an elicited-response low-inhibition task*. Paper presented at the biennial meeting of the Society for Research in Child Development, Montreal, Quebec, Canada.

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

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.

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

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

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.