Original Articles

Surprise! 20-month-old infants understand the emotional consequences of false beliefs

Rose M. Scott

School of Social Sciences, Humanities, and Arts, University of California Merced, 5200 N. Lake Road, Merced, CA 95343, United States

ABSTRACT

Recent studies suggest that by the second year of life, infants can attribute false beliefs to agents. However, prior studies have largely focused on infants' ability to predict a mistaken agent's physical actions on objects. The present research investigated whether 20-month-old infants could also reason about belief-based emotional displays. In Experiments 1 and 2, infants viewed an agent who shook two objects: one rattled and the other was silent. Infants expected the agent to express surprise at the silent object if she had a false belief that both objects rattled, but not if she was merely ignorant about the objects' properties. Experiment 3 replicated and extended these findings: if an agent falsely believed that two containers held toy bears (when only one did so), infants expected the agent to express surprise at the empty, but not the full, container. Together, these results provide the first evidence that infants in the second year of life understand the causal relationship between beliefs and emotional displays. These findings thus provide new evidence for false-belief understanding in infancy and suggest that infants, like older children, possess a robust understanding of belief that applies to a broad range of belief-based responses.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The ability to predict and interpret others' behavior in terms of their underlying mental states plays a vital role in everyday social interactions. Developmental psychologists have long been interested in when and how this important ability develops. In particular, considerable research has focused on when children first understand that other people can be mistaken, or hold false beliefs, about the world. Early investigations into this question used elicited-response tasks in which children had to answer direct questions about the behavior of an individual who held a false belief (e.g., Baron-Cohen, Leslie, & Frith, 1985; Gopnik & Astington, 1988; Perner, Leekam, & Wimmer, 1987; Wellman & Bartsch, 1988; Wimmer & Perner, 1983). The results of such tasks suggested that the ability to attribute false beliefs to others did not emerge until at least 4 years of age (for reviews, see Devine & Hughes, 2014; Wellman, Cross, & Watson, 2001). Recently, however, researchers have developed a number of alternative paradigms for assessing false-belief understanding in much younger children (e.g., Buttelmann, Carpenter, & Tomasello, 2009; Kovács, Téglás, & Endress, 2010; Luo, 2011; Onishi & Baillargeon, 2005; Scott, Baillargeon, Song, & Leslie, 2010; Southgate, Senju, &

Csibra, 2007). Positive results have now been obtained with infants aged 6–25 months using a variety of response measures (for reviews, see Baillargeon et al., 2015; Scott, Roby, & Smith, in press), leading many investigators to conclude that the capacity to attribute false belief to others emerges by at least the end of the first year of life (e.g., Baillargeon, Scott, & He, 2010; Barrett et al., 2013; Buttelmann et al., 2009; Carruthers, 2013; Kovács et al., 2010; Luo, 2011; Scott, in press; Southgate et al., 2007; Surian, Caldi, & Sperber, 2007).

However, several researchers have offered alternative accounts for these recent findings (e.g., Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; Heyes, 2014; Perner, 2010; Perner & Roessler, 2012; Ruffman, 2014). These accounts differ from one another in many respects, including the level of conceptual sophistication that they attribute to infants and the mechanisms that they assume underlie development (e.g., compare Butterfill & Apperly, 2013, to Heyes, 2014). Nevertheless, they share two common assumptions. First, they assume that infants' success in false-belief tasks does not reflect a genuine understanding of belief. Instead, they argue that infants' responses are driven by various alternative factors such as low-level perceptual novelty (e.g., Heyes, 2014), learned behavioral rules for how agents typically behave in particular situations (e.g., Ruffman, 2014), or an early-developing system for tracking belief-like states (e.g., Apperly &

Butterfill, 2009). Responses based on these factors sometimes coincide with the responses infants' would produce if they were tracking an agent's false belief, yielding what appears to be successful false-belief reasoning. Second, these accounts assume that infants' performance in false-belief tasks should exhibit sharp limits: they should fail tasks in which responses to these various alternative factors diverge from responses to beliefs.

One way to test these alternative accounts of infants' performance in false-belief tasks is thus to explore the range of situations in which infants succeed. When older children succeed in traditional elicited-response false-belief tasks, they do so in a variety of belief-inducing situations, such as those involving false beliefs about the location (e.g., Baron-Cohen et al., 1985), identity (e.g., Gopnik & Astington, 1988), or contents (e.g., Hogrefe, Wimmer, & Perner, 1986) of an object. They also demonstrate an understanding of a complex array of belief-based responses produced by a mistaken agent. For instance, they are able to predict and to explain (1) an agent's physical actions, such as where the agent searches, reaches, or points (e.g., Baron-Cohen et al., 1985; Low, 2010), (2) an agent's verbal behaviors, such as what the agent will say is in a particular container (e.g., Hogrefe et al., 1986), and (3) the agent's emotional responses when she discovers that her beliefs about a situation are false (e.g., Hadwin & Perner, 1991; Wellman & Banerjee, 1991). This ability to reason about a rich set of belief-based responses across a range of belief-inducing situations indicates that older children possess a robust understanding of false belief (e.g., Apperly, 2011; Low & Wang, 2011; Perner & Ruffman, 2005). If infants are capable of attributing false beliefs to agents, then they should be able to demonstrate the "flexible use of belief understanding" that is seen in older children (Perner & Ruffman, 2005, p. 216). If infants were instead limited to reasoning about an arbitrary subset of belief-inducing situations and belief-based responses, then this might suggest that infants' performance was driven by mechanisms other than an understanding of belief (e.g., Apperly & Butterfill, 2009; Heyes, 2014; Low & Watts, 2013; Perner & Roessler, 2012).

Are infants able to understand a range of belief-inducing situations and belief-based responses? A number of studies have now confirmed that infants can reason about different belief-inducing scenarios, including those involving false beliefs about the presence (e.g., Kampis, Parise, Csibra, & Kovács, 2015; Kovács et al., 2010; Southgate & Vernetti, 2014), location (e.g., Onishi & Baillargeon, 2005; Surian et al., 2007), identity (e.g., Buttelmann, Suhrke, & Buttelmann, 2015; Scott & Baillargeon, 2009; Scott, Richman, & Baillargeon, 2015; Song & Baillargeon, 2008), contents (e.g., Buttelmann, Over, Carpenter, & Tomasello, 2014), and properties (e.g., Scott et al., 2010) of objects. With regards to belief-based responses, however, studies have focused almost exclusively on infants' ability to predict and interpret a mistaken agent's physical actions on objects (e.g., Buttelmann et al., 2009; Luo, 2011; Scott & Baillargeon, 2009; Senju, Southgate, Snape, Leonard, & Csibra, 2011; Song & Baillargeon, 2008; Surian & Geraci, 2012; Träuble, Marinović, & Pauen, 2010; for an exception, see Southgate, Chevallier, & Csibra, 2010). It thus remains unclear whether infants understand the broad range of belief-based responses grasped by older children and, consequently, whether they possess a robust understanding of false belief.

In the present research, we sought to address this issue by examining infants' understanding of belief-based emotional responses. Specifically, *surprise* is a belief-based emotion: it occurs when one discovers that one's beliefs about a situation are false (e.g., Roseman, 2001; Roseman, Antoniou, & Jose, 1996). To illustrate, imagine that Sarah sees her mother place an apple in her lunchbox. Her mother later decides to replace the apple with a banana. If Sarah does not witness the swap and falsely believes that her lunchbox contains an apple, then she will be surprised

when she opens the lunchbox and discovers a banana. However, if she sees her mother make the exchange, then she will not be surprised when she sees the banana because surprise does not occur when one's beliefs are confirmed. Surprise also does not occur when one is merely ignorant and holds no particular expectation about a situation: if Sarah does not know which fruit her mother placed in the lunchbox, then she should not be surprised to find either a banana or an apple inside.[1] Do infants expect an agent to express surprise when (and only when) she discovers that her beliefs are false?

To date, children's understanding of the causal relationship between false belief and surprise has been investigated using elicited-response tasks in which children were asked direct questions about a mistaken agent's inner emotional state (e.g., how the agent would feel) or external emotional display (e.g., which object would cause the agent to make a surprised face) (e.g., Hadwin & Perner, 1991; MacLaren & Olson, 1993; Ruffman & Keenan, 1996; Wellman & Banerjee, 1991; Wellman & Bartsch, 1988). For instance, MacLaren & Olson (1993) tested children in a task adapted from the classic "Smarties" task (e.g., Gopnik & Astington, 1988; Hogrefe et al., 1986). Children were first shown a container that had unexpected contents. On some trials, the contents were desirable (e.g., a toothpaste container that held candy) and on others the contents were undesirable (e.g., a Smarties container that held rocks). This container was then placed alongside a second, visually identical container that held typical/expected contents. Children were asked which of these two containers would surprise a puppet that had not seen inside. 5- to 6-year-old children correctly indicated that the puppet would be surprised by the container with unexpected contents, irrespective of the contents' desirability. In contrast, 4-year-olds tended to select the container with desirable contents regardless of whether those contents were expected (e.g., a Smarties container with candy) or unexpected (e.g., a toothpaste container with candy). These results suggest that prior to age 5, children might incorrectly view surprise as resulting from something desirable.

Ruffman and Keenan (1996) argued that young children also incorrectly treat surprise as an ignorance-based rather than a belief-based emotion. In their task, children were introduced to two puppets, John and Katy, who wished to take one of two objects (e.g., a spoon or a bandaid) to their grandmother's house in a box. While John was outside, Katy placed one of the objects in the box (e.g., the spoon) and then left. In her absence, the experimenter replaced the contents of the box with the other object (e.g., the bandaid). Thus, Katy held a false belief about the contents of the box (she believed it was a spoon when it was a bandaid), while John was merely ignorant about which object was in the box (either a spoon or a bandaid). Both puppets returned and looked inside the box. Children were then asked which puppet felt surprised when they saw what was inside. Only the oldest children (7- to 8-year-olds) reliably chose the individual who held a false belief. Younger children (4- to 6-year-olds) chose randomly, suggesting that they viewed surprise as resulting from discovering something that one does not know (experienced by both puppets) rather than from discovering that one's beliefs are false.

Prior results from elicited-response tasks thus suggest that an understanding that surprise results specifically from discovering that one is mistaken may not emerge until middle childhood. However, as described above, considerable research has now shown

---

[1] Note that Sarah would be surprised if she were to find a snake in her lunchbox. Although Sarah holds no particular expectation about which fruit should be present (and hence should find neither surprising), she has a host of expectations about how the world typically works, including the kinds of things that mothers put in lunchboxes. Objects that violate these "background beliefs" about the world would be unexpected and elicit surprise (Ruffman & Keenan, 1996, p. 43).