

Running head: ATTRIBUTING FALSE BELIEFS ABOUT OBJECT IDENTITY

Which penguin is this?

Attributing false beliefs about object identity at 18 months

Rose M. Scott

and

Renée Baillargeon

University of Illinois at Urbana-Champaign

Please address correspondence to:

Rose Scott

Department of Psychology

University of Illinois

603 East Daniel Street

Champaign, IL 61820

Email: rscott2@uiuc.edu

Abstract

Recent research has shown that infants as young as 13 months can attribute false beliefs to agents, suggesting that the psychological-reasoning subsystem necessary for attributing reality-incongruent informational states (SS2) is operational in infancy. The present research asked whether 18-month-olds' false-belief reasoning extends to false beliefs about object identity. Infants watched events involving an agent and two toy penguins; one penguin could be disassembled (2-piece penguin) and the other could not (1-piece penguin). Infants realized that outdated contextual information could lead the agent to falsely believe she was facing the 1-piece rather than the 2-piece penguin, suggesting that 18-month-olds can attribute false beliefs about the identity of objects and providing new evidence for SS2 reasoning in the second year of life.

Psychologists have long been interested in determining at what age children become able to attribute false beliefs and false perceptions to others. Initial investigations suggested that this ability did not emerge until about 4 years of age (e.g., Flavell, 1988; Gopnik & Wellman, 1994; Perner, 1991). This evidence came primarily from tasks in which children were asked direct questions about an agent's beliefs or perceptions (e.g., Baron-Cohen, Leslie, & Frith, 1985; Gopnik & Astington, 1988; Wimmer & Perner, 1983). However, recent investigations conducted with tasks designed to tap children's *spontaneous* (as opposed to *elicited*) responses suggest that the ability to reason about false beliefs and false perceptions may emerge much earlier. Spontaneous-response tasks used to date include anticipatory-looking (AL) and violation-of-expectation (VOE) tasks; both are illustrated below.

Building on prior AL results with 3-year-olds (e.g., Clements & Perner, 1994; Garnham & Ruffman, 2001), Southgate, Senju, and Csibra (2007) tested 25-month-olds in a novel non-verbal AL task. The children first watched two videotaped familiarization events in which a bear puppet hid a toy in one of two boxes while a female agent looked on. The agent wore a visor and sat behind a panel with two small, closed doors, one above each box; only the agent's head was visible above the panel. The bear hid the toy in the left box in one familiarization event, and in the right box in the other event. After the bear hid the toy, the two doors lit up, and then the agent opened the correct door to retrieve the toy. During the test trial, the agent saw the bear hide the toy in the left box. At that point, a phone rang behind the agent, who turned toward the sound; while she was facing away, the bear removed the toy from the left box, briefly hid it in the right box, and then left with it. The phone then stopped ringing, the agent turned back toward the boxes, and the doors lit up. Most children correctly anticipated the agent's behavior and looked at the door above the left box, where she falsely believed the toy to be hidden. Similar results were found in another condition in which the agent falsely believed the toy to be in the right box. Together, these results suggest that, by 2 years of age, children already possess some understanding of false belief.

Findings from VOE tasks suggest that children younger than 2 years may also possess some

understanding of false belief. To date, there have been four reports using such tasks with infants aged 13 to 18 months; three of the reports examined infants' understanding of an agent's false belief about the location of an object (Onishi & Baillargeon, 2005; Song, Onishi, Baillargeon, & Fisher, 2008; Surian, Caldi, & Sperber, 2007), and one report also examined infants' understanding of an agent's false perception of an object (Song & Baillargeon, 2008). In this last report, 14.5-month-olds first watched live familiarization events in which a female agent faced two toys: a stuffed skunk and a doll with blue pigtails. Across trials, the toys were presented on placemats or inside shallow containers; the agent consistently reached for the doll, suggesting that she preferred it over the skunk. During the test trial, in the agent's absence (false-perception condition), the toys were hidden in boxes with lids: the doll in a plain box and the skunk in a box with a tuft of blue hair protruding from under its lid. The agent then returned and reached for either the plain box (plain-box event) or the hair box (hair-box event), and then paused. The infants who saw the plain-box event looked reliably longer than those who saw the hair-box event, suggesting that they expected the agent to falsely perceive the tuft of hair as belonging to the doll, and hence to falsely believe that the doll was hidden in the hair box and the skunk in the plain box. In another condition (true-perception condition), the agent witnessed the hiding of the doll and skunk. The infants who saw the hair-box event now looked reliably longer than those who saw the plain-box event, suggesting that they expected the agent to search for her doll where she had seen it being hidden. These conclusions were supported by the results of another experiment in which the agent consistently reached for the skunk instead of for the doll. The infants in the false-perception condition now expected the agent to reach for the plain box, whereas those in the true-perception condition expected her to reach for the hair box.

The goal of the present research was two-fold. First, in Experiments 1 and 2 we asked using a novel VOE task whether 18-month-old infants could attribute to an agent a false belief about the *identity* of an object. In their commentary on Onishi and Baillargeon (2005), Perner and Ruffman (2005) wrote: "The conclusions from the [standard] false-belief task are warranted only because understanding of false belief

around 4 years can be demonstrated in a variety of belief-inducing situations” (p. 216). We reasoned that evidence that infants can attribute to agents not only false beliefs about location and false perceptions, as discussed above, but also false beliefs about identity, would help demonstrate that infants, too, succeed “in a variety of belief-inducing situations”.

Second, in Experiment 3, we sought to address an alternative interpretation of the VOE results cited above (Onishi & Baillargeon, 2005; Song & Baillargeon, 2008; Song et al., 2008; Surian et al., 2007). According to this interpretation (e.g., Southgate et al., 2007), the responses of the infants in these tasks might reflect, not an ability to attribute false beliefs and false perceptions to others, but rather a general expectation that *ignorance leads to error* (henceforth, we refer to this as the error interpretation). As an example, consider the false-perception condition of Song and Baillargeon (2008). According to the error interpretation, the infants who saw the plain-box event could have looked reliably longer simply because (1) they realized that the agent was ignorant about the doll’s current location, the plain box; (2) they expected her to search for the doll in the wrong location, the hair box; and hence (3) they were surprised when she searched the correct location, the plain box. The error interpretation does not apply to Southgate et al. because the bear puppet left with the toy in the test trial so that *both* boxes were wrong locations. Had the infants simply expected the ignorant agent to search for the toy in a wrong location, they would have had no basis for predicting which box she would select; the fact that they expected her to search the box where she falsely believed the toy to be hidden suggests that they perceived the agent as mistaken, rather than as ignorant, about the toy’s location.

How one interprets the VOE results cited earlier as well as those reported here has critical implications for the age at which infants become able to attribute false beliefs and false perceptions (henceforth, false beliefs) to agents, and thus for the age at which the computational subsystem responsible for attributing these internal states comes online. To better explain these implications, we next present our account of early psychological reasoning (see also Luo & Baillargeon, 2007; Song & Baillargeon, 2008; Song et al., 2008) and then review findings from social neuroscience that appear consistent with this account.

Armed with these results, we then return to the present research.

An Account of Early Psychological Reasoning

Like several other researchers, we assume that infants are born with a psychological-reasoning system that provides them with a shallow causal framework for interpreting the intentional actions of agents (e.g., Gergely & Csibra, 2003; Johnson, 2005; Leslie, 1994; Premack & Premack, 1995; Surian et al., 2007). This assumption is by no means universal among infancy researchers; for alternative frameworks, see, for example, Meltzoff (2005), Tomasello, Carpenter, Call, Behne, and Moll (2005), and Woodward (2005).

Most system-based accounts share the following four assumptions, either explicitly or implicitly. First, the operation of the psychological-reasoning system is thought to be largely *unconscious*: infants are not aware of the causal framework they use when reasoning about agents, any more than young children are aware of the grammar of their language as they begin to understand and produce sentences (e.g., Leslie, 2000; Song et al., 2008). Second, the psychological-reasoning system is triggered when infants attempt to make sense of the intentional actions of *any* entity they construe as an agent—its activity is not restricted to human agents (e.g., Csibra, 2008; Kuhlmeier, Wynn, & Bloom, 2003; Luo & Baillargeon, 2005; Premack & Premack, 1997; Johnson, Shimizu, & Ok, 2007), or even to self-propelled agents (e.g., Wu & Baillargeon, 2007). Third, embedded in the psychological-reasoning system are a few core constraints or principles, including that of *rationality*: when pursuing a goal, agents are expected to select actions that are not only causally appropriate but also reasonably efficient (e.g., Csibra, 2008; Gergely & Csibra, 2003; Gergely, Bekkering, & Király, 2002). Finally, the psychological-reasoning system is often thought to consist of at least *two subsystems*, termed here Subsystem-1 (SS1) and Subsystem-2 (SS2); SS2 is assumed to come online after SS1 (e.g., Gergely & Csibra, 2003; Leslie, 1994; Tager-Flusberg, 2005).

Where our account of psychological reasoning begins to depart from prior system-based accounts is in how it defines the precise functions of SS1 and SS2, and therefore in how it characterizes the shift that takes place in infants' psychological reasoning as SS2 becomes operational. Most system-based accounts

assume that SS2 allows infants to attribute internal states to agents, but SS1 does not; instead, SS1 provides infants with a non-mentalistic, teleological system of action interpretation (e.g., Gergely & Csibra, 2003; Leslie, 1994). One key assumption of these accounts is that teleological reasoning is reality-based: a teleological infant should not be able to distinguish his representation of reality from that of an agent—*reality should be as construed by the infant*. However, contrary to this claim, recent evidence suggests that even young infants recognize that an agent's representation of a scene may not match their own (e.g., Liskowski, Carpenter, Striano, & Tomasello, 2006; Luo & Baillargeon, 2007; Luo & Johnson, 2009; Moll & Tomasello, 2004). In particular, infants who can see an object that is not visible to an agent use the agent's representation of the scene, rather than their own, to interpret the agent's responses. These results have led us to assume that *both* SS1 and SS2 allow infants to attribute internal states, albeit different ones, as explained below.

Subsystem-1

We assume that when infants watch an agent act on objects in a scene, SS1 allows them to attribute two kinds of internal states to the agent: *motivational* and *reality-congruent informational* states (see also Carpenter, Nagell, & Tomasello, 1998; Luo & Baillargeon, 2007; Premack, 1990). Motivational states specify the agent's motivation in the scene and include goals and dispositions. Reality-congruent informational states specify what knowledge or accurate information (as construed by the infant) the agent possesses about the scene. The agent's knowledge may come about through perception (i.e. what the agent can see), through memory (i.e. what the agent has seen), or through inference (i.e. what the agent can reasonably infer based on previous experiences in the scene or in the world more generally).

SS1 makes it possible for infants to represent not only states of *knowledge* but also states of *ignorance*. When critical information is missing from an agent's representation of a scene, so that this representation is *incomplete* relative to that of the infant (e.g., the agent does not see an object that the infant sees), SS1 allows the infant (1) to identify the missing information and (2) to reason about the agent's actions accordingly. The mechanism that is used to represent the agent's state of ignorance may be understood as a

masking mechanism: by masking the information about the scene that is not available to the agent, infants can interpret or predict the agent's actions in terms of the remaining, shared information.

Current evidence suggests that SS1 becomes operational in the first months of life and is well in place by the end of the first year. Thus, experiments on *motivational* states have revealed that even young infants can attribute dispositions and goals to agents (e.g., Csibra, 2008; Hamlin, Wynn, & Bloom, 2007; Luo & Baillargeon, 2005; Sommerville, Woodward, & Needham, 2005; Woodward, 1998). For example, after watching an agent repeatedly reach for object-A as opposed to object-B in a scene, 5-month-olds attribute to the agent a particular disposition, a preference for object-A over object-B. When the objects' positions are reversed, infants expect the agent to reach for object-A in its new position, and they look reliably longer if the agent reaches for object-B instead (e.g., Luo & Baillargeon, 2005; Woodward, 1998).

Similarly, experiments on *reality-congruent informational* states suggest that, by the end of the first year, infants (1) keep track of what objects an agent can or cannot see, and has or has not seen, in a scene and (2) use this information to interpret the agent's responses or to guide their own (e.g., Brooks & Meltzoff, 2005; Caron, Kiel, Dayton, & Butler, 2002; Liszkowski et al., 2006; Luo & Baillargeon, 2007; Tomasello & Haberl, 2003). Thus, 12.5-month-olds who watch a female agent repeatedly reach for object-A over object-B do *not* attribute to the agent a preference for object-A if object-B is hidden from her by a screen; however, they *do* attribute such a preference if the agent is aware of object-B's presence behind the screen because she saw it there earlier (Luo & Baillargeon, 2007). Recent experiments suggest that even 6-month-olds consider what objects an agent can or cannot see when interpreting her actions (Luo & Johnson, 2009).

Subsystem-2

We assume that SS2 extends SS1 in that it allows infants to attribute *reality-incongruent informational* states to agents. When an agent holds a *false* or a *pretend* belief about a scene (e.g., the agent believes that a toy is hidden in location-A but the infant knows it has been moved to location-B; the agent pretends that an empty cup is full of liquid), so that the agent's representation of the scene is *incompatible* with that of the

infant, SS2 allows the infant (1) to identify the agent's alternative beliefs about the scene and (2) to reason about the agent's actions accordingly (e.g., Leslie, 1987; Onishi, Baillargeon, & Leslie, 2007). The mechanism that is used to represent an agent's false or pretend beliefs (and other fictional internal states) is a *decoupling* mechanism (e.g., Leslie, 1987; Sommer et al., 2007). This mechanism enables infants to hold in mind two distinct versions of a scene: one that corresponds to reality (as they construe it), and one that incorporates the agent's false or pretend beliefs but otherwise functions as expected.

To illustrate the difference between the masking mechanism of SS1 and the decoupling mechanism of SS2, consider a situation where the infant knows that an agent's toy is hidden in location-B as opposed to location-A. If the agent does not know in which of the two locations the toy is hidden, SS1's masking mechanism can block out the information that the toy is in location-B; the infant will then have no expectation about which of the two locations the ignorant agent will search. In contrast, if the agent falsely believes that the toy is in location-A, SS1 is no longer sufficient: simply masking the information that the toy is in location-B is not enough to correctly predict the agent's actions. To do so, the infant must specify what belief the mistaken agent actually holds about the toy's hiding place. SS2's decoupling mechanism allows the infant to specify that the agent falsely believes the toy is in location-A, and thus to correctly predict that the agent will search in location-A.

At what point is SS2 operational? The answer to this question depends on how one interprets the VOE results mentioned earlier (Onishi & Baillargeon, 2005; Song & Baillargeon, 2008; Song et al., 2008; Surian et al., 2007). As the preceding example illustrates, our account argues that SS1 cannot explain infants' responses in these tasks: SS2 must also be involved. If infants expect a female agent who falsely believes her toy is in location-A to search location-A (and are surprised if she searches location-B instead), it cannot be because they attribute ignorance to the agent. According to our account, if infants attributed only ignorance to the agent, they would have no expectation about which of the two locations she would search. Instead, infants expect the agent to search at location-A because they represent her false belief that the toy is in location-A

and they expect her to act in accordance with this belief. Infants require SS2's decoupling mechanism to hold in mind the agent's alternative representation of the scene. Thus, from our perspective, the earliest evidence that SS2 is operational comes from the findings of Surian et al. (2007) with 13-month-olds.

Spontaneous- and Elicited-response Tasks

The account of early psychological reasoning we just presented leaves one important question unanswered: if infants at the end of the first year of life can already attribute reality-incongruent informational states to agents, as evidenced by their performance in spontaneous-response tasks, then why do children across countries typically fail elicited-response false-belief tasks until about age 4 (e.g., Callaghan et al., 2005; Liu, Wellman, Tardiff, & Sabbagh, 2008; Wellman, Cross, & Watson, 2001)? Previous explanations for this well-established finding have often appealed to some form of conceptual change: for example, it has been proposed that young children lack a concept of belief (e.g., Perner, 1991) or do not yet understand that beliefs are representations rather than copies of reality (e.g., Gopnik & Wellman, 1994). However, if infants can attribute false beliefs to agents in spontaneous-response tasks, such explanations are unlikely.

Our interpretation of young children's failure in elicited-response false-belief tasks takes a different approach. For ease of exposition, consider the classic Sally-Ann task (e.g., Baron-Cohen et al., 1985) in which children listen to the following story acted out with props: Sally faces a basket and a box; she hides a marble in the basket and then leaves; in her absence, Ann moves the marble to the box. Children are then asked where Sally will look for her marble when she returns. Beginning at about age 4, children typically answer correctly and point to the basket; prior to age 4, children typically point to the box, the marble's current location. We assume that success in the Sally-Ann task depends on the interaction of three separate processes. First, children must represent Sally's false belief about the marble's location; we assume that this process takes place in SS2 as children listen to the story (*false-belief-representation* process). Second, when asked the test question, children must attend to the question, decide to answer it, and tap their representation of Sally's false belief (*response-selection* process). Finally, children must inhibit any prepotent tendency to

answer the question based on their own knowledge of the marble's current location (*response-inhibition* process; e.g., Birch & Bloom, 2003; Carlson & Moses, 2001; Kovács, 2009; Leslie, German, & Polizzi, 2005). We believe that young children's difficulty with the Sally-Ann task lies not in the false-belief-representation process, as is often assumed, but rather in the response-selection and response-inhibition processes. Although children can and do represent Sally's false belief accurately, they have difficulty (1) tapping this representation when deciding how to answer the question and (2) inhibiting their tendency to respond based on their own knowledge of the marble's current location. Both of these difficulties are substantial: in false-belief tasks where little or no inhibition is required (e.g., Ann takes the marble away, so that children do not know its current location), young children typically perform at chance (e.g., Wellman et al., 2001).

The preceding interpretation clarifies why young children succeed at spontaneous-response tasks but fail at even low-inhibition elicited-response tasks. In spontaneous-response tasks, the SS2 *false-belief-representation* process is activated when children realize that the agent holds a false belief; the children often spontaneously reveal their understanding of this false belief in their reactions to the unfolding events (e.g., just as adults watching a movie often spontaneously reveal their understanding of the characters' thoughts and feelings). Low-inhibition elicited-responses tasks also activate the *response-selection* process because children are asked a test question. Children then fail because the joint activation of the false-belief-representation process and the response-selection process overwhelms their limited information-processing resources, and/or because the neural connections between the brain regions that serve these two processes are still immature and inefficient in early childhood (e.g. Johnson, 2001; Lebel, Walker, Leemans, Philips, & Beaulieu, 2008). Since young children typically succeed when asked questions about the agent's internal states that do not involve her false belief (e.g., "Did Sally see Ann move the marble?"), we must assume that the joint activation of SS1 and the response-selection process poses no problems for young children, and that the neural connections between the brain regions that serve these two processes are well established.

Full consideration of the evidence supporting the preceding interpretation is beyond the scope of this

article (e.g., Bunge, Hazeltine, Scanlon, Rose, & Gabrieli, 2002; Hester, D'Esposito, Cole, & Garavan, 2007; Rowe, Toni, Josephs, Frackowiak, & Passingham, 2000). Here we mention only one set of findings that may support our analysis of spontaneous- and elicited-response tasks. Research with adults by Waszak, Prinz, and their colleagues (e.g., Herwig, Prinz, & Waszak, 2007; Mueller, Brass, Waszak, & Prinz, 2007; Waszak et al., 2005) has revealed a critical distinction between actions that are triggered by external stimuli ("externally-guided actions") and actions that are generated by internal decision and selection processes ("internally-guided actions") (see also Obhi and Haggard, 2004). In externally-guided action tasks, adults simply respond to a stimulus (e.g., they press a left key if a stimulus appears on the left and a right key if the stimulus appears on the right); in internally-selected action tasks, adults decide what response to produce (i.e. they choose whether to press the left or the right key after the stimulus appears). Different brain regions are activated in the two types of tasks, even when participants' responses are equated to eliminate perceptual and motor confounds, suggesting that different brain circuits control internally- and externally-guided actions (Mueller et al., 2007). Specifically, internally-guided actions produce greater activation in the dorsal anterior cingulate cortex (ACC) and the right lateral prefrontal cortex (PFC) extending to the anterior PFC. Mueller et al. speculated that the activation in the ACC is related to "the internal selection of actions" (p. 1360), whereas the activation in the lateral and anterior PFC is related to the greater attentional and working-memory demands of internally-guided actions. One interesting possibility is that the distinction between spontaneous- and elicited-response false-belief tasks echoes that between externally- and internally-guided action tasks, respectively, and that young children's difficulties with elicited-response tasks reflect, at least in part, difficulties with the response-selection process associated with the ACC and the lateral and anterior PFC.

Why Two Subsystems?

Why do we posit two separate subsystems in infants' psychological-reasoning system? There are at least three reasons for doing so. First, as is hopefully clear from the previous sections, the masking mechanism of SS1 seems intuitively very different from the decoupling mechanism of SS2; blocking out a

portion of reality to interpret the actions of an ignorant agent seems computationally far simpler than holding in mind an alternative version of reality to interpret the actions of an agent with false or pretend beliefs.

Second, although autistic children succeed at a number of SS1 tasks (e.g., they are able to reason about what objects others can see; e.g., Hobson, 1984; Leslie & Frith, 1988), they appear to have specific difficulties with false-belief and pretense tasks, suggesting that their SS2 is selectively impaired (e.g., Baron-Cohen et al., 1985; Leslie & Frith, 1988; Peterson, Wellman, & Liu, 2005; Tager-Flusberg, 2005). Peterson et al. (2005) administered five psychological-reasoning tasks (including both SS1 and SS2 tasks) to a large sample of Australian children that included typically developing preschoolers, deaf native signers, deaf late signers, and children with autism. Results indicated that the last two groups of children were generally delayed relative to the first two. However, while the deaf late signers mastered all of the tasks in the same sequence as the typical preschoolers and deaf native signers, the autistic children did not: for them, the false-belief task represented the most challenging task in the sequence (for the other children, a task on a character's hidden emotion was the most challenging task). The authors concluded that "individuals with autism may have a distinctive, autism-specific difficulty with the sort of mental state understanding that is needed for false-belief tasks, above and beyond other sorts of mental state understandings" (p. 513).

Third, comparative findings indicate that non-human primates distinguish between intentional and accidental actions (e.g., Call & Tomasello, 1998; Call, Hare, Carpenter, & Tomasello, 2004), keep track of what objects others can or cannot see, and have or have not seen (e.g., Flombaum & Santos, 2005; Hare, Call, & Tomasello, 2001), and even deceive others by manipulating what they see or hear so as to induce a state of ignorance (e.g., Hare, Call, & Tomasello, 2006; Santos, Nissen, & Ferrugia, 2006). These various results suggest that non-human primates can represent both motivational and reality-congruent informational states. At the same time, there is no clear evidence yet that non-human primates can attribute false beliefs or comprehend pretense, suggesting that SS2 may have emerged later in evolution (e.g., Krachun, Carpenter, Call, & Tomasello, in press; Santos, Morticorena, & Goddu, 2007).

Supportive Evidence from Neuroscience

Findings from social neuroscience are beginning to shed light on the nature and organization of the brain regions that are selectively engaged during psychological reasoning in adults (for developmental reviews, see Diamond, 2002; Saxe, Carey, & Kanwisher, 2004a). Do some of these findings appear consistent with the system-based account of early psychological reasoning we presented in the last section? There are, of course, enormous differences in the richness and sophistication of adults' and infants' psychological-reasoning abilities; thus, we fully acknowledge from the outset the considerable interpretive difficulties in comparing adults' and infants' responses in psychological-reasoning tasks, and in extrapolating from findings about adult brains to infant brains. We merely hope that the following speculations may help identify possible similarities between adults' and infants' psychological reasoning and suggest additional links that might be explored in future research.

Our system-based account of early psychological reasoning makes at least four suggestions about the neural underpinnings of psychological reasoning in adults; all of these suggestions have some supportive evidence. First, psychological-reasoning tasks should engage different brain regions than tasks that do not require reasoning about the intentional actions of agents, such as tasks that involve reasoning about physical events or about outdated photographs (e.g., Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006; Saxe & Kanwisher, 2003). Second, similar brain regions should be activated when subjects reason about the intentional actions of either human or non-human agents (e.g., Castelli, Frith, Happé, & Frith, 2002; Castelli, Happé, Frith, & Frith, 2000; Gallagher et al., 2000). Third, focusing on SS1 alone, it should be difficult to separate the brain regions involved in reasoning about agents' motivational as opposed to reality-congruent informational states, because these must go hand in hand: in order to act on objects in a scene, an agent must detect, remember, or infer the presence and location of the objects (e.g., Jellema, Baker, Wicker, & Perrett, 2000; Pelphrey, Singerman, Allison, & McCarthy, 2003; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004b). Fourth, and most relevant to the present research, on the assumption that the brain regions that

serve SS1 and SS2 are somewhat distinct, we would expect a substantial, but not a complete, overlap between the brain regions activated during true-belief (or knowledge) and false-belief tasks.

To see why, consider once again the doll experiment described earlier (Song & Baillargeon, 2008). To respond correctly in the test trial, the infants in the true-perception condition had to reason that (1) the agent preferred the doll over the skunk; (2) this preference should lead her to form the goal of searching for the doll; (3) the agent could see the plain and the hair box; and (4) the agent knew that the doll had been hidden in the plain box. According to our account, these steps all involved motivational and reality-congruent informational states and as such would have activated the brain regions associated with SS1; the regions associated with SS2 would not have been activated. In the false-perception condition, the infants had to reason that (1) the agent preferred the doll over the skunk; (2) this preference should lead her to form the goal of searching for the doll; (3) the agent could see the plain and the hair box; and (4) the agent could infer, based on the familiarization trials, that the doll and skunk were both present and hidden in the boxes. These steps all involved SS1, leading one to expect substantial overlap in the brain regions activated in the true- and false-perception conditions. However, in the false-perception condition, additional steps were needed to represent the agent's reality-incongruent informational states: the infants had to reason that (5) the agent was likely to *falsely perceive* the tuft of hair protruding from the hair box as part of the doll; and hence (6) the agent was likely to *falsely believe* that the doll was hidden in the hair box. According to our account, only SS2 could have allowed the infants to specify the agent's false perception and false belief, and to form an expectation about where she was likely to search given her erroneous informational states.

Consistent with our account, a recent fMRI experiment found significant activation differences between false-belief and true-belief trials, suggesting that the brain regions associated with false- and true-belief reasoning do not fully overlap. Sommer et al. (2007) presented adults with false- and true-belief stories adapted from the Sally-Ann task described earlier (Baron-Cohen et al., 1985). Each story consisted of seven pictures organized in three sections. In the first four pictures (baseline section), agent-1 hid an

object in location-A while agent-2 watched; agent-1 then left, and agent-2 removed the object from location-A. In the next two pictures (belief-induction section), agent-2 moved the object to location-B, either before (false-belief story) or after (true-belief story) agent-1 returned. In the final picture (outcome section), agent-1 searched for the object in either location-A or location-B. When contrasting the belief-induction section of the false-belief and true-belief trials, the authors found increased activation in the right temporo-parietal junction (TPJ-R) as well as in regions of the frontal cortex including the dorsal ACC, the right dorsolateral PFC, and the right lateral anterior PFC. The TPJ-R was specifically activated during the false-belief trials, leading the authors to conclude that their results pointed to “the role of the TPJ-R in the decoupling mechanism” and more generally “in computing mental states that create a perspective difference, such as a person’s false belief that contrasts with the state of reality” (p. 1383; see also Perner et al., 2006).

If we return to our analysis of the processes involved in elicited-response false-belief tasks (i.e., the false-belief-representation, response-selection, and response-inhibition processes), the results of Sommer et al. (2007) suggest that the TPJ-R plays an important role in the representation of false beliefs, which we assume is carried out by SS2’s decoupling mechanism (see also Kobayashi, Glover, & Temple, 2007; Sabbagh, Bowman, Evraire, & Ito, this volume; Saxe & Wexler, 2005). In addition, comparison of the results of Sommer et al. with those Mueller et al. (2007) obtained in their internally-selected action task reveals tantalizing parallels: both experiments found significant activations in the ACC and the lateral and anterior PFC, suggesting that these brain regions play a role in the response-selection process. We return to these issues in the General Discussion.

The Present Research

The evidence reviewed in the preceding sections suggests that (1) the ability to represent reality-incongruent informational states, such as false beliefs or pretense, may depend on a separate subsystem of the psychological-reasoning system, SS2; (2) the brain regions that are selectively engaged during SS2 reasoning may include the TPJ; and (3) SS2 may already be operational in the second year of life, pointing

to the early maturation of at least some of the brain regions associated with SS2. To date, the evidence for this last conclusion has come from VOE tasks examining whether infants can attribute to an agent a false perception of an object (Song & Baillargeon, 2008) or a false belief about an object's location (Onishi & Baillargeon, 2005; Song et al., 2008; Surian et al., 2007). The present research sought to extend these results and examined 18-month-olds' ability to attribute to an agent a false belief about an object's *identity*.

Although the ability to attribute a false perception is related to the ability to attribute a false belief about identity, the two are nevertheless distinct. A false perception involves coming to an erroneous conclusion about which *type of object* one is facing based on current, misleading perceptual information. For instance, an agent might falsely perceive a deceptive candle as an apple or a deceptive tree branch as a snake. A false belief about identity involves coming to an erroneous conclusion about which particular *object token* one is facing based on contextual information; this contextual information was once helpful in identifying tokens correctly, but is now outdated and therefore misleading. To illustrate, imagine a scenario in which a girl and her brother are playing with identical toys; whenever she leaves the scene, the girl stores her toy in a special box. The next time she is away, her brother replaces the toy in the box with his own, identical toy. When the girl returns, she does not have a false perception of the toy in the box—she can determine at a glance what type of object it is—but she does have a false belief about its identity. The available contextual information—the fact that the toy is in the box—suggests to her that it is her toy, when in fact it is not. Thus, the contextual information that initially helped the girl identify her toy is now outdated, leading her to hold a false belief about the identity of the toy in the box.

In Experiment 1, infants watched an agent interact with two identical toy penguins. One of the penguins could be disassembled into two pieces (2-piece penguin) and one could not (1-piece penguin); when assembled, the 2-piece penguin was visually indistinguishable from the 1-piece penguin. Thus, as with the example above, the agent needed to use contextual information (rather than perceptual information) to determine whether a particular penguin was the 1-piece penguin or the assembled 2-piece

penguin. If the infants understood that contextual cues could lead the agent to conclude that she was facing the 1-piece penguin when she was really facing the assembled 2-piece penguin, then they would be attributing to the agent a false belief about the identity of the penguin. Experiment 2 sought to confirm and clarify the results of Experiment 1, and Experiment 3 tested whether the results of Experiments 1 and 2 could be explained more parsimoniously by an error interpretation (e.g., Southgate et al., 2007).

Experiment 1

The infants in Experiment 1 were assigned to a false- or a true-belief condition. The infants in both conditions received the same four familiarization trials (see Fig. 1). In the first two trials, the agent sat at a large window in the back wall of the apparatus, behind two small platforms; the 1-piece penguin stood in front of one platform, and the disassembled 2-piece penguin stood in front of the other platform. An experimenter's gloved hand placed the 1-piece penguin on its platform, placed the two pieces of the 2-piece penguin on the other platform, and then exited the apparatus. The agent then placed a metal key in the bottom piece of the 2-piece penguin, stacked the top piece on the bottom piece, and then paused until the trial ended. The next two familiarization trials were identical except that the platforms were replaced with shallow containers and the penguins' locations were reversed. The familiarization trials thus served to establish that (1) the 1- and 2-piece penguins were present in each trial; (2) the 1-piece penguin did not come apart (as demonstrated by the fact that it remained in one piece when the experimenter lifted it by the top of its head); (3) the 2-piece penguin was disassembled at the start of each trial; (4) the two penguins could appear in different locations (left or right) and arrangements (on platforms or inside shallow containers); (5) the agent had the goal of hiding her key; and finally (6) the agent consistently sought the 2-piece as opposed to the (otherwise identical) 1-piece penguin because only the 2-piece penguin provided a hiding location for her key.

Next, the infants received two test trials. In the *false-belief* condition (see Fig. 2), the agent was absent at the start of each trial: her window was closed. The gloved hand assembled the 2-piece penguin and covered it with a large transparent cover; the hand then covered the 1-piece penguin with a large opaque

cover (either cover was wide enough to hold the two pieces of the 2-piece penguin side by side). After the hand exited the apparatus, the agent entered the apparatus with her key, reached for either the transparent (transparent-cover event) or the opaque (opaque-cover event) cover, and paused until the trial ended. The test trials in the *true-belief* condition (see Fig. 3) were similar except that the agent was present throughout the trials and thus knew that the 2-piece penguin was under the transparent cover.

To succeed at the present task, the infants in the false-belief condition had to attribute to the agent a complex, interlocking set of internal states. First, the agent's actions during the familiarization trials suggested that she had a goal of hiding her key in the 2-piece penguin, which she was likely to maintain during the test trials. Second, when entering the apparatus in the test trials, the agent should correctly infer, based on the familiarization trials, that the 1- and 2-piece penguins were both present. Third, the agent should falsely expect, again based on the familiarization trials, that the 2-piece penguin was disassembled. Fourth, this expectation should lead the agent to form the false belief that the penguin under the transparent cover was the 1-piece penguin. Fifth, the agent's false belief about the identity of the penguin under the transparent cover should lead her to falsely believe that the disassembled 2-piece penguin was under the opaque cover. If the infants attributed to the agent this set of motivational states, reality-congruent informational states, and reality-incongruent informational states, then they should expect her to reach for the opaque cover, and they should be surprised when she reached for the transparent cover instead. The infants should thus look reliably longer at the transparent- than at the opaque-cover event.

Success in the true-belief condition required fewer state attributions. First, as in the false-belief condition, the infants had to attribute to the agent the goal of hiding her key in the 2-piece penguin, and they had to expect this goal to be maintained in the test trials. Second, since the agent had watched while the gloved hand assembled the 2-piece penguin and placed the transparent cover over it, the infants should expect her to reach for that cover, and they should be surprised when she reached for the opaque one instead. The infants should thus look reliably longer at the opaque- than at the transparent-cover event.

Method

Participants

Participants were 28 healthy term infants, 14 male and 14 female (ages 17 months, 6 days to 18 months, 6 days, $M = 17$ months, 22 days). Another 8 infants were tested but excluded, 4 because they looked the maximum time allowed on both test trials, 2 because the difference in their test looking times was over 2.5 standard deviations from the mean of their condition, and 2 because they were overly active (1) or required a bathroom break (1). Half the infants were randomly assigned to the false-belief condition ($M = 17$ months, 22 days), and half to the true-belief condition ($M = 17$ months, 22 days).

The infants' names in this and the following experiments were obtained primarily from purchased mailing lists and birth announcements in the newspaper. Parents were reimbursed for their transportation but were not compensated for participation. The racial and ethnic composition of the infants tested was 81% Caucasian, 6% Asian (or mixed Asian-Caucasian), 6% Hispanic (or mixed Hispanic-Caucasian), 4% African American (or mixed African American-Caucasian), and 1% American Indian; 2% of the parents chose "other race." No information was collected on parents' education, occupation, and income.

Apparatus

The apparatus consisted of a wooden display booth (46.5 cm high X 102 cm wide X 57 cm deep), mounted 78 cm above the room floor. The infant faced a large opening (45 cm X 95 cm) in the front of the apparatus; between trials, a curtain consisting of a muslin-covered wooden frame (59.5 cm X 101.5 cm) was lowered in front of this opening. The side walls of the apparatus were painted white, the back wall was made of white foam board, and the floor was covered with pale green contact paper.

The agent wore a green shirt and sat on a wooden chair centered behind a large window (46 cm X 51 cm) in the back wall of the apparatus. This back window extended from the apparatus floor, was located 7 cm from the right wall, and had two doors that could be closed to hide the agent; each door (46 cm X 25.5 cm) was made of white foam board. In each trial, the agent held a thin silver metal key 5 cm long and 2 cm

wide (at its widest point). A large muslin screen behind the agent hid the test room.

The experimenter wore a long white glove on her right hand and arm and sat behind a window (51 cm X 38 cm) in the right wall of the apparatus. This right window was filled with a white fringe curtain and was located 4 cm above the apparatus floor and 7 cm from the back wall.

In the first two familiarization trials, two platforms (each 4.5 cm X 18 cm X 11.5 cm) stood 5 cm apart on the apparatus floor, 16.5 cm in front of the back window; the right platform was 5.5 cm from the right window. Each platform was made of foam board 0.5 cm thick and was covered with red contact paper. In the last two familiarization trials, the platforms were replaced with two shallow containers (each 3 cm X 18 cm X 11.5 cm). Each container was made of foam board 0.5 cm thick and was covered inside and out with yellow contact paper. At the start of the test trials, two covers, one transparent and one opaque, stood 5 cm apart on the apparatus floor, 4.5 cm in front of the back window. The transparent cover was on the left and the opaque cover was on the right, 5.5 cm from the right window. Each cover (14.5 cm X 18 cm X 11.5 cm) had a wooden spherical knob, 3 cm in diameter, affixed to the center of its top surface. The transparent cover was made of clear Plexiglas 0.3 cm thick; the opaque cover was made of foam board 0.5 cm thick and was covered with a white and brown patterned contact paper.

Although the 1- and 2-piece penguins were both designed to separate into a top and a bottom piece, the 1-piece penguin was glued together so that it no longer came apart. The top piece of each penguin was made of white and light blue plastic and included its head and stomach; a row of small, white, fuzzy pompoms was glued to the lower edge of the top piece, around the front. The bottom piece of each penguin consisted of a cylinder resting on a light blue plastic base. The cylinder was covered with dark blue contact paper and decorated with colorful dots; the base sported three small, yellow, fuzzy pompons on either side at the front. The 1-piece penguin was 12 cm high, 9 cm wide, and 8 cm deep (at the largest points). The bottom piece of the 2-piece penguin was closed 0.5 cm below its top edge, providing a surface on which the agent could place her key. The inside of the bottom piece was covered with the same dark

blue contact paper as the outside. The top piece of the 2-piece penguin rested on a thin circular base (hidden by the white pompoms) that fit easily inside the bottom piece; when the top piece was stacked on the bottom piece, the 2-piece penguin was indistinguishable from the 1-piece penguin.

The two penguins stood on the apparatus floor at the start of every trial. In the first two familiarization trials, the 1-piece penguin stood centered 1 cm in front of one platform. The top and bottom pieces of the 2-piece penguin stood next to each other 1 cm in front of the other platform; the top piece was on the left, with its left edge aligned with that of the platform, and the bottom piece was on the right, with its right edge aligned with that of the platform. Similar arrangements were used with the shallow containers in the last two familiarization trials; in the test trials, the covers were placed further back to start, so that the penguins rested 13 cm in front of the covers.

The infants were tested in a brightly lit room; three fluorescent tubes attached to the front and back of the apparatus provided additional light. Two frames, each 183 cm high and 76 cm wide and covered with blue cloth, stood at an angle on either side of the apparatus and isolated the infants from the test room.

Trials

Each trial consisted of an initial and a final phase; looking times during the two phases were computed separately. During the *initial* phase of a trial, the experimenter and agent performed the scripted actions appropriate for the trial, ending with a paused scene; during the *final* phase, the infant watched this paused scene until the trial ended. The duration of the initial phase was fixed and depended on the specific actions performed. The duration of the final phase was infant-controlled (see below for the specific criteria used to end trials). When a trial ended, a supervisor lowered the curtain at the front of the apparatus; inter-trial intervals lasted about 10 s, and each new trial began with the raising of the curtain.

In the following descriptions, the numbers in parentheses indicate the number of seconds taken to perform the actions described. To help the experimenter and agent adhere to the events' scripts, a metronome beat softly once per second. A camera mounted behind the infant captured the events, and a

second camera mounted beneath the apparatus floor captured the infant; the two images were combined using a mixer, projected onto a TV behind the apparatus, and recorded onto a computer. During the test session, the supervisor monitored the events on the TV to confirm that they followed the prescribed scripts.

Familiarization trials. Each familiarization trial consisted of a 25-s initial phase followed by a final phase. At the start of each trial, the agent sat at the back window with her hands in her lap and her eyes focused on a neutral point on the apparatus floor between the two platforms or containers. While acting on objects (e.g., placing her key inside the bottom piece of the 2-piece penguin), the agent kept her eyes on the objects; otherwise, she kept her eyes on the neutral point and did not make eye contact with the infant.

The two penguins stood in front of the two platforms at the start of the first two familiarization trials, and in front of the two containers at the start of the last two familiarization trials. For half the infants in each experimental condition (left-right side condition), the two pieces of the 2-piece penguin stood in front of the left platform in the first two trials (left familiarization trials) and in front of the right container in the last two trials (right familiarization trials); the agent thus reached to different locations for the 2-piece penguin in the left and right trials. For the other infants (right-left condition), these locations were reversed, so that the infants received right then left familiarization trials. For ease of description, the familiarization trials are described from the perspective of the infants in the left-right condition (see Fig. 1).

During the initial phase of the left familiarization trials, the experimenter's gloved right hand entered the apparatus through the right window (1 s), grasped the top of the 1-piece penguin (1 s), and placed it on the right end of the right platform (3 s). Next, the hand grasped the bottom piece of the 2-piece penguin (2 s) and placed it on the right end of the left platform (2 s); the hand then grasped the top piece of the 2-piece penguin (2 s), placed it on the left end of the left platform (2 s), and exited the apparatus (2 s). After a pause (1 s), the agent brought both hands into the apparatus (1 s); she placed her left hand on the floor 18.5 cm behind the platforms, and she held the key in her right hand 20 cm above the floor, 5 cm to the left of and 3 cm behind the left platform. Next, the agent tilted the key up and down (2 s), to call attention to it.

She then grasped the bottom piece of the 2-piece penguin with her left hand (1 s) and placed the key inside the bottom piece with her right hand (2 s). The agent then grasped the top piece of the 2-piece penguin with her right hand (1 s) and placed it on top of the bottom piece (2 s), so that the 2-piece penguin now looked identical to the 1-piece penguin. The agent then paused, with her right hand holding the top piece of the 2-piece penguin and her left hand holding the bottom piece. During the final phase, the infants watched this paused scene until the computer signaled that the trial had ended.

The right familiarization trials were identical to the left familiarization trials except that the shallow containers were substituted for the platforms, the locations of the 1- and 2-piece penguins were reversed, and the gloved hand moved the pieces of the 2-piece penguin before moving the 1-piece penguin. The gloved hand always manipulated the penguin on the right prior to the penguin on the left.

Test trials. During the test trials, the infants saw the transparent- and opaque-cover events appropriate for their experimental condition. Each trial consisted of a 33-s initial phase followed by a final phase. At the start of each trial in the *false-belief condition* (see Fig. 2), the agent was absent and the doors in the back window were closed. The two pieces of the 2-piece penguin stood in front of the transparent cover, on the left, and the 1-piece penguin stood in front of the opaque cover, on the right. During the initial phase of each trial, the gloved hand entered the apparatus (1 s), grasped the 1-piece penguin (1 s), and moved it back to a dot centered 6 cm in front of the opaque cover (3 s). Next, the hand grasped the bottom piece of the 2-piece penguin (2 s) and moved it back to a dot centered 6 cm in front of the transparent cover (2 s); the hand then grasped the top piece of the 2-piece penguin (2 s) and stacked it on the bottom piece (2 s), so that the 2-piece penguin now looked identical to the 1-piece penguin. The hand then grasped the knob of the transparent cover (1 s), lifted it approximately 15 cm (1 s), moved it forward over the assembled 2-piece penguin (1 s), and lowered it to the floor (2 s). Next, the hand performed the same actions with the opaque cover and lowered it over the 1-piece penguin (5 s). The hand then withdrew from the apparatus (1 s). After a pause (1 s), the agent opened the left (1 s) and then the right (1 s) door in the

back window, placed her hands in her lap (1 s), and paused (1 s), looking at a neutral point on the apparatus floor between the covers. She then brought her hands into the apparatus (1 s), placing her left hand on the floor behind the covers, and holding the key in her right hand in the same position as in the familiarization trials. She then tilted the key up and down (2 s), grasped the knob of either the transparent (transparent-cover event) or the opaque (opaque-cover event) cover with her left hand (1 s), and paused. During the final phase of each test trial, the infants watched this paused scene until the trial ended.

The test trials in the *true-belief condition* (see Fig. 3) were identical to those in the false-belief condition with one exception: the agent was present throughout both trials. At the start of each test trial, the agent sat at the back window with her hands in her lap and her eyes focused on the neutral point between the two covers. The gloved hand performed the same actions as in the false-belief condition and then exited the apparatus. The agent paused for 5 s (since she did not have to open the doors in the back window), then brought her hands into the apparatus (1 s), tilted the key (2 s), grasped the knob of either the transparent (transparent-cover event) or the opaque (opaque-cover event) cover with her left hand (1 s), and paused. During the final phase, the infants watched this paused scene until the trial ended.

Procedure

The infant sat on a parent's lap centered 45 cm in front of the apparatus. Parents were instructed to remain silent and neutral and to close their eyes during the test trials. Prior to the start of the test session, the agent showed the infant her key for about 5 s.

The infants' looking behavior was monitored by two naïve observers who viewed the infant through peepholes in the cloth-covered frames on either side of the apparatus. Each observer held a button linked to a computer and pressed the button when the infant attended to the event. The looking times recorded by the primary observer were used to determine when a trial had ended.

The infants in the false- and true-belief conditions received the same four familiarization trials. Half the infants in each condition received two left familiarization trials first, and half received two right

familiarization trials first. Examination of the looking times during the 25-s initial phase at the start of each familiarization trial indicated that the infants in both the false-belief ($M = 23.9$) and the true-belief ($M = 24.1$) conditions were highly attentive. The final phase of each familiarization trial ended when the infants either (1) looked away for 2 consecutive seconds after having looking for at least 5 cumulative seconds or (2) looked a total of 60 cumulative seconds without looking away for 2 consecutive seconds.

Next, the infants received two test trials in which they saw the transparent- and opaque-cover events appropriate for their condition. In each condition, half the infants saw the transparent-cover event first and half saw the opaque-cover event first. Examination of the looking times during the 33-s initial phase at the start of each test trial indicated that the infants in both the false-belief ($M = 32.5$) and the true-belief ($M = 30.4$) conditions were highly attentive. The final phase of each test trial ended when the infants (1) looked away for 2 consecutive seconds after having looked for at least 4 cumulative seconds or (2) looked a total of 40 cumulative seconds without looking away for 2 consecutive seconds.

To assess interobserver agreement during the familiarization and test trials, each trial was divided into 100-ms intervals, and the computer determined in each interval whether the observers agreed on whether or not the infant was looking at the event. Percent agreement was calculated for each trial by dividing the number of intervals in which the observers agreed by the total number of intervals in the trial. Interobserver agreement was measured for 23/28 infants (only one observer was present for the other infants) and averaged 94% per trial per infant.

Preliminary analyses of the test data in this and the following experiments revealed no interactions of condition and event with either sex or order, all $F_s < 1.07$; the data were therefore collapsed across sex and order in subsequent analyses.

Results

The infants' looking times during the final phases of the two left and the two right *familiarization* trials were averaged and analyzed by means of a 2 X 2 analysis of variance (ANOVA) with condition (false-

or true-belief) as a between-subjects factor and side (left or right) as a within-subject factor. The analysis revealed no significant main effect of condition, $F(1, 26) = 0.66$, or side, $F(1, 26) = 0.04$, and no significant interaction between these two factors, $F(1, 26) = 0.05$. The infants in the false- and true-belief conditions thus tended to look equally whether the agent reached for the 2-piece penguin on the left or on the right in the familiarization trials (false-belief: left $M = 18.2$, $SD = 8.8$, right $M = 19.6$, $SD = 12.9$; true-belief: left $M = 21.0$, $SD = 11.6$, right $M = 20.9$, $SD = 8.7$).

The infants' looking times during the final phase of each *test* trial (see Fig. 4) were analyzed using a 2 X 2 ANOVA with condition (false- or true-belief) as a between-subjects factor and event (transparent- or opaque-cover) as a within-subject factor. The analysis yielded a significant interaction between condition and event, $F(1, 26) = 22.01$, $p < .0001$. No other effect was significant, both $F_s(1, 26) < 1.30$, $p_s > .266$. Planned comparisons revealed that the infants in the false-belief condition looked reliably longer at the transparent-cover ($M = 25.5$, $SD = 9.5$) than at the opaque-cover ($M = 18.8$, $SD = 10.7$) event, $F(1, 26) = 13.17$, $p = .001$, whereas the infants in the true-belief condition looked reliably longer at the opaque-cover ($M = 21.0$, $SD = 10.2$) than at the transparent-cover ($M = 15.4$, $SD = 8.6$) event, $F(1, 26) = 9.03$, $p = .006$. Non-parametric analyses confirmed these results. In the false-belief condition, 12/14 infants looked longer at the transparent-cover event, Wilcoxon signed-ranks $T = 19$, $p = .035$; in the true-belief condition, 11/14 infants looked longer at the opaque-cover event, $T = 6$, $p = .002$.

Discussion

The infants in the true-belief condition looked reliably longer at the opaque- than at the transparent-cover event, suggesting that they (1) attributed to the agent, during the familiarization trials, the goal of hiding her key in the 2-piece penguin; (2) expected this goal to be maintained during the test trials and to lead the agent to again seek the 2-piece penguin; and hence (3) expected the agent to reach for the transparent cover, where she had seen the 2-piece penguin being placed, and were surprised when she reached for the opaque cover instead.

Unlike the infants in the true-belief condition, the infants in the false-belief condition looked reliably longer at the transparent- than at the opaque-cover event. This result suggested that the infants (1) attributed to the agent, during the familiarization trials, the goal of hiding her key in the 2-piece penguin; (2) expected this goal to be maintained in the test trials and to lead the agent to again seek the 2-piece penguin; (3) understood that the agent would correctly infer, based on the familiarization trials, that the 1- and 2-piece penguins were both present in the apparatus; (4) realized that the agent would falsely expect, based on the familiarization trials, that the 2-piece penguin was disassembled; (5) reasoned that this expectation would lead the agent to form the false belief that the penguin under the transparent cover was the 1-piece penguin; (6) further reasoned that the agent's false belief about the identity of the penguin under the transparent cover would lead her to falsely believe that the disassembled 2-piece penguin was under the opaque cover; and (7) expected the agent to reach for the opaque cover and were surprised when she reached for the transparent cover instead. The result of the false-belief condition thus suggested that, by 18 months of age, infants can already attribute a false belief about an object's identity to an agent.

According to the preceding interpretation, the infants in the false-belief condition attributed to the agent *two* false beliefs: one about the identity of the penguin visible under the transparent cover (the agent assumed it was the 1-piece penguin when it was really the 2-piece penguin), and one about the location of the 2-piece penguin (the agent assumed it was under the opaque cover when it was really under the transparent cover). Could the infants have attributed to the agent *only* a false belief about the location of the 2-piece penguin? We think this interpretation is unlikely. In the present research, the 2-piece penguin was not hidden in one location in the agent's presence and then moved to a new location in her absence. If the agent falsely believed that the 2-piece penguin was under the opaque cover, it could only be because she *first* falsely concluded that the penguin visible under the transparent cover was the 1-piece penguin.

Experiment 2

The primary goal of Experiment 2 was to replicate the result of the false-belief condition in

Experiment 1. A secondary goal was to validate one portion of our interpretation of the infants' responses in the true- and false-belief conditions. We have suggested that the infants attributed a complex, interlocking set of internal states to the agent, which included the goal of hiding her key in the 2-piece penguin. However, it might be argued that the infants could have ignored or dismissed the key and simply attributed to the agent a particular disposition, a *preference* for the 2- as opposed to the 1-piece a penguin. After all, the agent held on to the 2-piece penguin throughout the paused scene at the end of each familiarization trial, so such a preference might seem plausible (e.g., Luo & Beck, in press; Woodward, 1998).

The infants in Experiment 2 were assigned to a false-belief condition identical to that in Experiment 1 or to a new no-key condition; this condition was similar to the false-belief condition except that the agent performed her actions *without a key*. We predicted that, as in Experiment 1, the infants in the false-belief condition would look reliably longer at the transparent- than at the opaque-cover test event. As for the no-key condition, we reasoned that, if the infants in Experiment 1 attributed to the agent a simple preference for the 2-piece penguin, then the results of the no-key condition should be identical to those of the false-belief condition. On the other hand, if the infants in Experiment 1 attributed to the agent the goal of hiding her key in the 2-piece penguin, then the results of the no-key condition might differ from those of the false-belief condition. If the infants in the no-key condition viewed the two penguins as interchangeable, then they should look about equally whether the agent reached for the transparent or the opaque cover: either penguin should be acceptable to her. Such a negative result would not only underscore the importance of the agent's goal in the false-belief conditions of Experiments 1 and 2, but also help rule out low-level interpretations of these conditions, since the no-key condition was identical in all other respects.

Method

Participants

Participants were 28 healthy term infants, 15 male and 13 female (ages 17 months, 6 days to 18 months, 13 days, $M = 17$ months, 20 days). Another 10 infants were excluded, 7 because they were fussy

(4), distracted (2), or overly active (1), and 3 because the difference in their test looking times was over 2.5 standard deviations from the mean of their condition. Half the infants were randomly assigned to the false-belief condition ($M = 17$ months, 24 days), and half to the no-key condition ($M = 17$ months, 16 days).

Apparatus, Trials, and Procedure

The apparatus, trials, and procedure in Experiment 2 were identical to those in the false-belief condition of Experiment 1, except that in the no-key condition the infants were not shown the key prior to the start of the test session and the agent performed her actions without a key. In the familiarization trials, once the gloved hand had exited the apparatus, the agent placed her left hand on the apparatus floor while holding her right hand with its palm towards the infant. She tilted her hand up and down for 2 s, grasped the bottom piece of the 2-piece penguin with her left hand (1 s), touched the interior of the bottom piece with her right hand (1 s), and finally grasped the top piece of the 2-piece penguin with her right hand (1 s) and placed it on top of the bottom piece (2 s). The agent then paused, with her right hand holding the top piece of the 2-piece penguin and her left hand holding the bottom piece. As in the false-belief condition, the initial phase of each familiarization trial lasted about 25 s and ended with the same paused scene. In the test trials, once the gloved hand had withdrawn from the apparatus, the agent opened the doors in the back window (3 s), placed her hands in her lap (1 s), and paused (1 s), looking at a neutral point on the apparatus floor between the covers. She then brought her hands into the apparatus (1 s), placing her left hand on the floor behind the covers and holding her right hand in the same position as in the familiarization trials. She then tilted her hand up and down (2 s), grasped the knob of either the transparent (transparent-cover event) or the opaque (opaque-cover event) cover with her left hand (1 s), and paused. As in the false-belief condition, the initial phase of each test trial lasted about 33 s and ended with a similar paused scene, except for the fact that the agent's right hand held no key.

As in Experiment 1, half the infants in each condition received two left and then two right familiarization trials and half received two right and then two left familiarization trials. Furthermore, in each

condition, half the infants saw the transparent-cover event first in the test trials and half saw the opaque-cover event first. The infants in both conditions were highly attentive during the initial phases of the familiarization (false-belief: $M = 23.8$; no-key: $M = 23.3$) and test (false-belief: $M = 32.0$; no-key: $M = 32.0$) trials. Interobserver agreement was calculated for 20/28 infants and averaged 94% per trial per infant.

Results

The infants' looking times during the final phases of the *familiarization* trials were averaged and analyzed by means of a 2 X 2 ANOVA with condition (false-belief or no-key) as a between-subjects factor and side (left or right) as a within-subject factor. The analysis revealed no significant main effect of condition, $F(1, 26) = 0.00$, or side, $F(1, 26) = 0.04$, and no interaction between these factors, $F(1, 26) = 0.00$. The infants in the false-belief and no-key conditions thus tended to look equally whether the agent reached for the 2-piece penguin on the left or on the right in the familiarization trials (false-belief: left $M = 20.5$, $SD = 11.4$, right $M = 21.0$, $SD = 14.4$; no-key: left $M = 20.6$, $SD = 9.0$, right $M = 21.2$, $SD = 14.0$).

The infants' looking times during the final phase of each *test* trial (see Fig. 4) were analyzed using a 2 X 2 ANOVA with condition (false-belief or no-key) as a between-subjects factor and event (transparent- or opaque-cover) as a within-subject factor. The analysis revealed a main effect of condition, $F(1, 26) = 12.58$, $p = .002$, and a significant interaction between condition and event, $F(1, 26) = 5.57$, $p = .026$. Planned comparisons indicated that the infants in the false-belief condition looked reliably longer at the transparent-cover ($M = 28.8$, $SD = 10.4$) than at the opaque-cover ($M = 18.4$, $SD = 8.9$) event, $F(1, 26) = 7.83$, $p = .010$, whereas the infants in the no-key condition looked equally at the transparent-cover ($M = 13.2$, $SD = 7.7$) and opaque-cover ($M = 15.2$, $SD = 12.0$) events, $F(1, 26) = 0.29$. Non-parametric analyses confirmed these results. In the false-belief condition, 11/14 infants looked longer at the transparent-cover event, $T = 14$, $p = .013$; in the no-key condition, only 7/14 infants did so, $T = 49$, $p > .20$.

Discussion

The infants in the false-belief condition looked reliably longer at the transparent- than at the

opaque-cover event. This result replicates that found in Experiment 1 and provides further evidence that 18-month-olds can attribute to an agent a false belief about an object's identity.

In contrast to the infants in the false-belief condition, the infants in the no-key condition tended to look equally at the two test events. This result supports the claim that the infants in the false-belief conditions of Experiments 1 and 2 and in the true-belief condition of Experiment 1 attributed to the agent, not a simple preference for the 2- over the 1-piece penguin, but a specific goal of hiding her key in the 2-piece penguin. When the agent had no key, the infants perceived the two penguins as interchangeable—a plausible stance, since the penguins were indistinguishable when the 2-piece penguin was assembled—and they looked about equally when the agent reached for the transparent or the opaque cover because lifting either cover would allow her to retrieve a penguin.

Experiment 3

We have argued that the infants in the false-belief conditions of Experiments 1 and 2 looked reliably longer at the transparent- than at the opaque-cover event because they attributed to the agent (1) a false belief about the identity of the penguin visible under the transparent cover and hence (2) a false belief about the location of the 2-piece penguin. As we discussed in the Introduction, an alternative interpretation of the present results, and of similar VOE results (Onishi & Baillargeon, 2005; Song & Baillargeon, 2008; Song et al., 2008; Surian et al., 2007), might be that infants possess a general expectation that *ignorance leads to error* (Southgate et al., 2007). According to this error interpretation, the infants in the false-belief conditions (1) attributed to the agent the goal of hiding her key in the 2-piece penguin; (2) expected this goal to lead the agent to again seek the 2-piece penguin in the test trials; (3) expected the agent to infer, based on the familiarization trials, that the 1- and 2-piece penguins were both present; (4) recognized that the agent was ignorant about the 2-piece penguin's current location (the transparent cover); and hence (5) expected the agent to *search in the wrong location* (the opaque cover). In this view, the infants looked reliably longer at the transparent- than at the opaque-cover event simply because they expected her to

search in the wrong location and were surprised when she searched in the correct location instead.

According to the account of early psychological reasoning we presented in the Introduction, SS1 alone could *not* explain the responses of the infants in the false-belief conditions of Experiments 1 and 2. When infants know that an object is hidden in location-A, but the agent does not know whether it is hidden in location-A or in location-B, SS1's masking mechanism allows infants to represent the agent's state of ignorance and leads them to have no expectation about which of the two locations the agent will search. Thus, the fact that the infants in the false-belief conditions of Experiments 1 and 2 expected the agent to reach for the opaque cover (and were surprised when she reached for the transparent cover instead) suggests that SS2's decoupling mechanism was also involved and enabled the infants (1) to represent the agent's false beliefs and (2) to reason about where she was likely to search given her false beliefs. In contrast, the error interpretation presented above assumes that SS1 alone *could* explain the infants' responses as long as they *also* possessed a general expectation that ignorance leads to error.

Is it plausible that infants in the second year of life would possess a general expectation that ignorance leads to error? In daily life, ignorance typically leads to random rather than to incorrect behavior. When one looks for a spoon in an unfamiliar kitchen, one may find the spoon on the first try or only after opening several drawers. When an ignorant agent searches for an object in one of two locations, do infants expect the agent to err, or do they hold no expectation about which location she will search? Experiment 3 was designed to examine this question. The infants were assigned to an *ignorance* condition similar to the false-belief conditions of Experiments 1 and 2, with one exception: instead of one transparent and one opaque cover, the test events now involved two identical covers (see Fig. 5). Half the infants saw two transparent covers (transparent-covers condition), and half saw two opaque covers (opaque-covers condition). In the transparent-covers condition, when the agent opened the doors in the test trials, the assembled 2-piece penguin was under the left transparent cover, and the 1-piece penguin was under the right transparent cover; because the two penguins were identical, the agent could not determine which

penguin was under which cover. In the opaque-covers conditions, the assembled 2-piece penguin was under the left opaque cover, and the 1-piece penguin was under the right opaque cover. Although the agent could infer that the 1- and 2-piece penguins were under the covers (since they had always been present in previous trials), she again had no basis for determining which penguin was under which cover.

We reasoned that if infants *do* expect ignorance to lead to error, then the infants in Experiment 3 (1) should recognize that the agent was ignorant about whether the 2-piece penguin was under the left or the right cover; (2) should expect her to search the incorrect cover (the one on the right); and (3) should be surprised when she reached for the correct cover instead (the one on the left). The infants should thus look reliably longer at the correct- than at the incorrect-cover event. On the other hand, if infants do *not* expect ignorance to lead to error, then the infants in Experiment 3 (1) should recognize that the agent was ignorant about whether the 2-piece penguin was under the left or the right cover and (2) should hold no expectation about which cover she would search. The infants should thus look about equally at the two test events.

Method

Participants

Participants were 16 healthy term infants, 8 male and 8 female (ages 17 months, 6 days to 18 months, 5 days, $M = 17$ months, 21 days). One additional infant was excluded because of parental intervention. Half the infants were randomly assigned to the transparent-covers condition ($M = 17$ months, 16 days), and half to the opaque-covers condition ($M = 17$ months, 25 days),

Apparatus, Trials, and Procedure

The apparatus, trials, and procedure in Experiment 3 were identical to those in the false-belief conditions of Experiments 1 and 2, with two exceptions. First, all infants received two left and then two right familiarization trials (Experiments 1 and 2 showed no order effects for the familiarization trials, so this variable was removed from Experiment 3). Second, the test events involved two identical transparent (transparent-covers condition) or opaque (opaque-covers condition) covers. In one (correct-cover) event,

the agent reached for the cover over the 2-piece penguin, which corresponded to the transparent cover in the previous experiments; in the other (incorrect-cover) event, she reached for the cover over the 1-piece penguin, which corresponded to the opaque cover in the previous experiments. In each condition, half the infants saw the correct-cover event first, and half saw the incorrect-cover event first.

The infants were highly attentive during the initial phases of the familiarization ($M = 24.4$) and test trials ($M = 32.3$). Interobserver agreement was calculated for 12/16 infants and averaged 93% per trial per infant. Preliminary analyses of the test data revealed no significant differences between the transparent- and opaque-covers conditions, $F(1, 14) = 0.00$, and no significant interaction between cover condition and event, $F(1, 14) = 0.27$ (transparent-covers: correct $M = 20.7$, $SD = 11.5$; incorrect $M = 24.4$, $SD = 7.8$; opaque-covers: correct $M = 23.0$, $SD = 12.7$; incorrect $M = 22.1$, $SD = 11.5$). The data were therefore collapsed across cover condition in subsequent analyses.

Results

The infants' looking times were compared to those of the infants in the combined false-belief conditions of Experiments 1 and 2.

The infants' looking times during the final phases of the *familiarization* trials were averaged and analyzed using a 2 X 2 ANOVA with condition (false-belief or ignorance) as a between-subjects factor and side (left or right) as a within-subject factor. The analysis revealed no significant effect of condition, $F(1, 42) = 1.44$, $p = .237$, or side, $F(1, 42) = 0.02$, and no interaction between these factors, $F(1, 42) = 0.28$. The infants in the false-belief and ignorance conditions thus looked equally when the agent reached for the 2-piece penguin on the left or on the right in the familiarization trials (false-belief: left $M = 19.4$, $SD = 10.0$, right $M = 20.3$, $SD = 13.5$; ignorance: left $M = 24.0$, $SD = 11.7$, right $M = 22.3$, $SD = 12.3$).

The infants' looking times during the final phase of each *test* trial (see Fig. 4) were analyzed using a 2 X 2 ANOVA with condition (false-belief or ignorance) as a between-subjects factor and event (correct/transparent- or incorrect/opaque-cover) as a within-subject factor. The analysis yielded a significant

interaction between condition and event, $F(1, 42) = 5.35, p = .026$. No other effect was significant, both $F_s(1, 42) < 2.80, p_s > .102$. Planned comparisons revealed that, whereas the infants in the combined false-belief conditions looked reliably longer at the transparent-cover ($M = 27.1, SD = 9.9$) than at the opaque-cover ($M = 18.6, SD = 9.7$) event, $F(1, 42) = 10.90, p = .002$, those in the ignorance condition looked about equally at the correct-cover ($M = 21.9, SD = 11.8$) and incorrect-cover ($M = 23.2, SD = 9.5$) events, $F(1, 42) = 0.16$. Finally, whereas 23/28 infants in the combined false-belief conditions looked longer at the transparent- than at the opaque-cover event, $T = 63, p = .001$, 10/16 infants in the ignorance condition looked longer at the correct- than at the incorrect-cover event, $T = 68, p > .20$.

Further Results

The results of Experiment 3 suggested that the infants (1) realized that the agent could not know under which cover the 2-piece penguin was located and (2) had no expectation about which cover she would reach for. This interpretation casts doubt on the idea that infants have a general expectation that ignorance leads to error. However, an alternative interpretation of the results was that the infants did have such an expectation, but could not demonstrate it because they had difficulty remembering which penguin was under which cover. This interpretation is unlikely given the results of the true-belief condition in Experiment 1: the infants could only have succeeded if they remembered (and assumed the agent also remembered) which penguin was under which cover. Nevertheless, to provide direct evidence against this alternative interpretation, 18-month-olds were tested in a true-belief condition identical to the ignorance condition of Experiment 3 except that the agent was present throughout the trials.

Participants were 12 infants, 6 male and 6 female (ages 17 months, 7 days to 17 months, 27 days, $M = 17$ months, 18 days). Another 4 infants were excluded because they were fussy (1), distracted (1), or unwilling to continue (1), or because the difference in the infant's test looking times was over 2.5 standard deviations from the mean of the condition (1). In the test trials, half the infants saw opaque covers and half saw transparent covers; half the infants saw the correct-cover event first and half saw the incorrect-cover

event first. The infants looked reliably longer at the incorrect-cover ($M = 21.6$, $SD = 8.3$) than at the correct-cover ($M = 12.6$, $SD = 7.9$) event, $F(1, 22) = 7.35$, $p = .013$, with 10/12 infants showing this pattern, $T = 9.5$, $p = .018$. The infants thus remembered that the 2-piece penguin was under the left cover, expected the agent to reach for that cover, and were surprised when she reached for the other cover instead. This result provides evidence that the infants in the ignorance condition looked equally at the correct- and incorrect-cover events because (1) they realized that the agent could not know the location of the 2-piece penguin and hence (2) they had no expectation about which cover she would reach for.

Discussion

According to the error interpretation proposed by Southgate et al. (2007), the results of the false-belief conditions in Experiments 1 and 2, and those of previous VOE false-belief tasks (Onishi & Baillargeon, 2005; Song & Baillargeon, 2008; Song et al., 2008; Surian et al., 2007), might simply reflect infants' general expectation that ignorance leads to error. Experiment 3 did not support this interpretation. The assembled 2-piece penguin and the 1-piece penguin were now placed under identical covers, either transparent or opaque. When the agent witnessed these actions, the infants expected her to reach for the correct cover, the one over the 2-piece penguin; when the agent did not witness these actions, the infants did not expect her to reach for the incorrect cover. Contrary to the error interpretation, the infants appeared to have no expectation as to which of the two covers the agent would reach for.

Although the results of Experiment 3 rule out the error interpretation proposed by Southgate et al. (2007), it could be argued that the present results, and those of previous VOE false-belief tasks (Onishi & Baillargeon, 2005; Song & Baillargeon, 2008; Song et al., 2008; Surian et al., 2007), are still open to alternative interpretations that grant infants the ability to represent states of ignorance but not false beliefs. Here we consider once such interpretation, by Wellman (in press); for a discussion of additional alternative interpretations, see Song and Baillargeon (2008) and Song et al. (2008).

According to Wellman (in press), infants might look reliably longer when an agent who holds a false

belief about the location of an object searches its current location, not because they are surprised that she is not acting in accordance with her false belief, but because (1) they realize that the agent is ignorant about the object's current location; (2) they have no expectation about which location she will search; and (3) they are surprised when she approaches the correct location directly or confidently, as opposed to tentatively or hesitatingly, as might be expected if the agent were truly ignorant. In Wellman's words, "why should she do that when she was unaware where it was?" We refer to this interpretation as the *uncertainty* interpretation, to distinguish it from the error interpretation of Southgate et al. (2007).

The results of Experiment 3 help rule out this uncertainty interpretation. If infants simply showed surprise when an ignorant agent reached unhesitatingly for an object's current location, then the infants in the false-belief conditions of Experiments 1 and 2 and in the ignorance condition of Experiment 3 should have responded similarly and looked reliably longer when the ignorant agent reached unhesitatingly for the 2-piece penguin's current location. However, this was not the case: In contrast to the infants in the false-belief conditions of Experiments 1 and 2, those in the ignorance condition of Experiment 3 looked about equally when the agent reached for the correct or the incorrect cover, even though she approached both covers with equal confidence (the agent's actions in reaching for a cover were exactly the same in all of the present experiments). This result thus casts doubt on the notion that infants in VOE false-belief tasks are merely surprised when an ignorant agent reaches unhesitatingly for an object's hiding place.

General Discussion

According to the account of early psychological reasoning we presented in the Introduction, infants' psychological-reasoning system consists of at least two subsystems, SS1 and SS2. When watching an agent act on objects in a scene, SS1 specifies the agent's *motivational* states (e.g., the agent's goal) and *reality-congruent informational* states (e.g., what knowledge the agent possesses or lacks about the scene). SS2 extends SS1 by specifying the agent's *reality-incongruent informational* states (e.g., what false or pretend beliefs the agent holds about the scene). Although many researchers agree that SS2 is likely to

come online after SS1, exactly when it does so has been the subject of long-standing debate (e.g., Gopnik & Wellman, 1994; Leslie, 1987, 2000; Perner, 1991). Studies using *elicited-response* false-belief tasks (which measure children's answers to direct questions about an agent's false belief) suggest that SS2 is not operational until about age 4 (e.g., Callaghan et al., 2005; Liu et al., 2008; Wellman et al., 2001). In contrast, studies using *spontaneous-response* false-belief tasks (which measure children's spontaneous responses as they reason about an agent's false belief) point to a much earlier age of onset. AL tasks have shown that 25-month-olds correctly anticipate the actions of an agent who holds a false belief about an object's location (Southgate et al., 2007); and VOE tasks have shown that 13- to 18-month-olds look reliably longer when an agent fails to act in accordance with her false perception of an object (Song & Baillargeon, 2008) or her false belief about an object's location (Onishi & Baillargeon, 2005; Song et al., 2008; Surian et al., 2007). The present research extends these results by showing that 18-month-olds also look reliably longer when an agent fails to act in accordance with her false belief about an object's *identity*.

The present results are also important in that they make clear that infants reason about agents' internal states "causally and systematically" (Saxe et al., 2004a, p. 91). To arrive at a correct expectation about the agent's actions during the test trials, the infants in the false-belief conditions of Experiments 1 and 2 had to reason that (1) the agent had the goal of hiding her key in the 2-piece penguin; (2) the agent would infer, based on the familiarization trials, that the 1- and 2-piece penguins were both present and that the 2-piece penguin was disassembled; (3) this last inference would lead the agent to form the false belief that the penguin visible under the transparent cover was the 1-piece penguin; and (4) this false belief would in turn lead the agent to falsely believe that the 2-piece penguin was hidden under the opaque cover. Success in the false-belief conditions thus required the attribution of a complex, interlocking set of motivational states, reality-congruent informational states, and reality-incongruent informational states.

Finally, the present results also cast doubt on two alternative interpretations of the present and previous VOE false-belief tasks: the error interpretation (Southgate et al., 2007) and the uncertainty

interpretation (Wellman, in press). The results of the ignorance condition in Experiment 3 make clear that infants (1) do not possess a general expectation that ignorance leads to error and (2) do not simply show surprise when an ignorant agent reaches unhesitatingly for an object's hiding place. When an agent does not know in which of two locations an object is hidden, infants appear to hold no specific expectation as to which of the two locations the agent will search.

Implications from Social Neuroscience

In the Introduction, we suggested that infants succeed at *spontaneous-response* false-belief tasks, such as the one used in the present research, because (1) these tasks involve primarily a false-belief-representation process and (2) by the end of the first year, SS1 and SS2 in infants' psychological-reasoning system are sufficiently mature to carry out this process, at least under optimal conditions. Thus, when infants watch events such as those shown in the false-belief conditions of Experiments 1 and 2, SS1 specifies the agent's motivational and reality-congruent informational states and SS2 specifies her reality-incongruent informational states, allowing the infants to form an expectation about the agent's actions. We speculated that the brain regions that serve SS1 and SS2 are somewhat distinct, and we reported findings by Sommer et al. (2007) consistent with this prediction: when contrasting false-belief trials with true-belief trials in an fMRI experiment, these authors found increased activation in the right TPJ (and other brain regions). Other researchers have also suggested that the TPJ is one of the key brain regions associated with false-belief reasoning. For example, Saxe and Kanwisher (2003) presented adults with false belief stories such as the following: "John told Emily that he had a Porsche. Actually, his car is a Ford. Emily doesn't know anything about cars." (p. 1841). Participants were then asked whether Emily, when she saw John's car, would think it was a Porsche or a Ford. When false-belief stories were compared to stories about outdated photographs (i.e. stories in which photographs depicted situations that were no longer true), the authors found increased activation in the TPJ bilaterally.

We also suggested that young children tend to fail at even *low-inhibition elicited-response* false-

belief tasks because these tasks involve both a false-belief-representation and a response-selection process (a third process, response-inhibition, is also involved in high-inhibition elicited-response false-belief tasks). In our analysis, children fail at low-inhibition elicited-response false-belief tasks because (1) the joint activation of the false-belief-representation process and the response-selection process overwhelms their limited attentional and working-memory resources, and/or (2) the neural connections between the brain regions that serve the two processes are still immature and inefficient during early childhood. We speculated that the response-selection process is related to the brain circuit Mueller et al. (2007) identified when comparing adults' responses in internally- as opposed to externally-guided action tasks: regions in the ACC and in the lateral and anterior PFC were differentially activated by subjects' internally guided actions. Interestingly, Sommer et al. (2007) also reported greater activation in these same areas in their (elicited-response) false-belief trials.

Forward and Backward Reasoning about False Beliefs

In this final section, we discuss a new direction for developmental research that is suggested by findings in adult neuroscience. These findings indicate that the superior temporal sulcus (STS) may also be involved in SS2 reasoning (e.g., Grèzes, Frith, & Passingham, 2004; Pelphrey, Morris, & McCarthy, 2004, 2005). For example, Grèzes et al. (2004) showed adults videotaped events in which an agent walked to a box on the floor of a room and picked it up; on some trials, the agent had been told the true weight of the box, and on other trials, the wrong weight. Participants judged, based on the agent's postural readjustments when lifting the box, whether the agent had a true or a false expectation about the weight of the box. In one analysis focusing exclusively on trials in which the agent's expectation was false, the authors contrasted trials in which participants judged correctly (and thus attributed a false expectation to the agent) with trials in which they judged incorrectly (and thus attributed a true expectation to the agent). Activations were found in the STS (and other brain regions). Pelphrey et al. (2005) showed normal and autistic adults computer-animated vignettes in which a target appeared in one of six locations in the visual

field of a human-like female agent; the agent then shifted her eyes either toward the target (congruent trials) or toward one of the empty locations (incongruent trials). Although both the normal and the autistic participants showed activation of the right STS region during the task, only the normal participants showed greater activation during the incongruent as opposed to the congruent trials.

Why do some tasks evoke more activation in the TPJ, and others in the STS? One critical difference between these tasks might be the following. In TPJ-activation tasks, participants establish that an agent *has* a false belief and then use this information to form an expectation about what she will do next. In STS-activation tasks, in contrast, participants infer, upon observing an agent's ineffectual or incorrect actions, that she *must have had* a false belief. To put it differently, whereas TPJ-activation tasks require *forward reasoning* about an agent's false belief (i.e., given the agent's false belief, what will she do next?), STS-activation tasks require *backward reasoning* (i.e., given the agent's actions, could she have had a false belief?). The increased activation of the STS in the latter case might signal that participants (1) are positing a false belief (a speculation that autistic participants may be less likely to produce) or more generally (2) are scrutinizing the agent's actions (if still visible) or going over them in their minds (if not) in an attempt to infer her likely informational and motivational states. According to this second possibility, STS activation might thus arise in both SS1 and SS2 tasks whenever subjects speculate about the internal states that could have led an agent to act as she did. Findings from Saxe et al. (2004b) are consistent with this view. Adults watched videotaped events in which a human agent passed behind a large bookcase as she walked across a room; on some (short-occlusion) trials, the agent did not stop behind the bookcase, but on other (long-occlusion) trials she stopped for a few seconds. When contrasting the long-occlusion trials with the short-occlusion trials, the authors found significantly more activation in a region of the right posterior STS. The authors speculated that this region might "have been recruited for revision of the original action representation, when the action did not unfold as expected" (p. 1444).

The preceding analysis suggests that both the TPJ and the STS may be associated with SS2, with

the TPJ being primarily engaged during forward false-belief reasoning tasks and the STS during backward false-belief reasoning tasks. This analysis suggests an interesting direction for infancy research. To date, VOE and AL false-belief tasks have all been tasks (like TPJ-activation tasks) in which infants are asked to reason about the actions of an agent who holds a false belief about a situation. However, it should also be possible to examine, as in STS-activation tasks, whether infants can *infer* from an agent's actions that her expectations about a situation were false. As an example, consider a condition similar to the false-belief conditions of Experiments 1 and 2, with the following exceptions: when the curtain opens at the start of the first test trial, the scene is already set, with the transparent and opaque covers already in place over the penguins. The agent enters the apparatus; at this point, both the infant and the agent should assume, based on the familiarization trials, that the 1-piece penguin is under the transparent cover and that the two pieces of the 2-piece penguin are standing side by side under the opaque cover. If the agent then peeks under the opaque cover (in a way that does not reveal what is under it to the infant) and looks hugely surprised, infants may infer that something else is under the opaque cover. If the agent then lifts the cover, infants should look reliably longer if the two pieces of the 2-piece penguin are revealed than if only the top piece, or no piece, or some other object is revealed. Finding that infants can infer, through backward reasoning, that an agent held a false expectation about an object would provide converging evidence that the ability to reason about reality-incongruent informational states is well in place in the second year of life.

Acknowledgments

This research was supported by a predoctoral traineeship from NIMH (1 T32MH1819990) to Rose Scott and by a grant from NICHD to Renée Baillargeon (HD-021104). We thank Cindy Fisher, Henry Wellman, and two anonymous reviewers for helpful comments on our manuscript; Gergo Csibra, Janice Juraska, Veronika Krieghoff, Sarah Laszlo, Alan Leslie, and Florian Waszak for helpful discussions; the staff of the University of Illinois Infant Cognition Laboratory for their help with the data collection; and the parents and infants who participated in the research.

References

- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*, 37-46.
- Birch, S. A. J., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science*, *14*, 283-286.
- Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*, 535-543.
- Bunge, S. A., Hazeltine, E., Scanlon, M. D., Rosen, A. C., & Gabrieli, J. D. E. (2002). Dissociable contributions of prefrontal and parietal cortices to response selection. *NeuroImage*, *17*, 1562-1571.
- Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). 'Unwilling' versus 'unable': Chimpanzees' understanding of human intentional action. *Developmental Science*, *7*, 488-498.
- Call, J., & Tomasello, M. (1998). Distinguishing intentional action from accidental action in Orangutans, Chimpanzees, and human children. *Journal of Comparative Psychology*, *112*, 192-206.
- Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., et al. (2005). Synchrony in the onset of mental-state reasoning. *Psychological Science*, *16*, 378-384.
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032-1053.
- Caron, A. J., Kiel, E. J., Dayton, M., & Butler, S. (2002). Comprehension of the referential intent of looking and pointing between 12 and 15 months. *Journal of Cognition and Development*, *3*, 445-464.
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *255*, 1-143.
- Castelli, F., Frith, C., Happe, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, *125*, 1839-49.

- Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, *12*, 314-325.
- Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377-395.
- Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, *107*, 705-717.
- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. Stuss & R. Knight (Eds.), *Principles of Frontal Lobe Function* (pp. 466-503). New York: Oxford University Press.
- Flavell, J. H. (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 244-267). New York: Cambridge University Press.
- Flombaum, J. I. & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology*, *15*, 447-452.
- Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C.D. (2000). Reading the mind in cartoons and stories: An fMRI study of "theory of mind" in verbal and nonverbal tasks. *Neuropsychologia*, *38*, 11-21.
- Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: Is anticipatory looking really related to understanding of belief? *Developmental Science*, *4*, 94-100.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Science*, *7*, 287-292.
- Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*, 6873.
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*, 26-37.
- Gopnik, A., & Wellman, A. (1994). The "theory theory". In L. Hirschfield & S. Gelman (Eds.), *Domain*

- specificity in culture and cognition* (pp. 257-293). New York: Cambridge University Press.
- Grèzes, J., Frith, C. D., & Passingham, R. E. (2004). Inferring false beliefs from the actions of oneself and others: An fMRI study. *NeuroImage*, *21*, 744-750.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*, 557-559.
- Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behavior*, *61*, 139-151.
- Hare, B., Call, J., & Tomasello, M. (2006). Chimpanzees deceive a human competitor by hiding. *Cognition*, *101*, 495-514.
- Herwig, A., Prinz, W., & Waszak, F. (2007). Two modes of sensorimotor integration in intention-based and stimulus-based action. *The Quarterly Journal of Experimental Psychology*, *60*, 1540-1554.
- Hester, R., D'Esposito, M., Cole, M. W., & Garavan, H. (2007). Neural mechanisms for response selection: Comparing selection of responses and items from working memory. *NeuroImage*, *34*, 446-454.
- Hobson, R. P. (1984). Early childhood autism and the question of egocentrism. *Journal of Autism and Developmental Disorders*, *14*, 85-104.
- Jellema, T., Baker, C. I., Wicker, B., & Perrett, D.I. (2000). Neural representation for the perception of the intentionality of actions. *Brain and Cognition*, *44*, 280-302.
- Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience*, *2*, 475-483.
- Johnson, S. C. (2005). Reasoning about intentionality in preverbal infants. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The Innate Mind: Structure and Contents* (pp. 254-271). New York: Oxford University Press.
- Johnson, S. C., Shimizu, Y. A., & Ok, S. (2007). Actors and actions: The role of agent behavior in infants' attribution of goals. *Cognitive Development*, *22*, 310-322.
- Kobayashi, C., Glover, G. H., & Temple, E. (2007). Children's and adults' neural bases of verbal and

- nonverbal 'theory of mind.' *Neuropsychologia*, 45, 1522-1532.
- Kovács, Á. M. (2009) Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, 12, 48-54.
- Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (in press). A competitive nonverbal false belief task for children and apes. *Developmental Science*.
- Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, 14, 402-408.
- Lebel, C., Walker, L., Leemans, A., Phillips, L., & Beaulieu, C. (2008). Microstructural maturation of the human brain from childhood to adulthood. *NeuroImage*, 40, 1044-1055.
- Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review*, 94, 412-426.
- Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119-148). New York: Cambridge University Press.
- Leslie, A. M. (2000). How to acquire a 'representational theory of mind'. In D. Sperber (Ed.), *Metarepresentations: An Multidisciplinary perspective* (pp. 197-223). Oxford: Oxford University Press.
- Leslie, A.M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, 6, 315-324.
- Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology*, 50, 45-85.
- Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, 7, 173-187.
- Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A., (2008). Theory of mind development in Chinese

- children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, 44, 523-531.
- Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, 16, 601-608.
- Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, 105, 489-512.
- Luo, Y., & Beck, W. (in press). Do you see what I see? Infants' reasoning about others' incomplete perceptions. *Developmental Science*.
- Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, 12, 142-149.
- Meltzoff, A. N. (2005). Imitation and other minds: The "like me" hypothesis. In S. Hurley & N. Chater (Eds.), *Perspectives on Imitation: From neuroscience to social science* (Vol. 2, pp. 55-77). Cambridge, MA: MIT Press.
- Moll, H., & Tomasello, M. (2004). 12- and 18-month-olds follow gaze behind barriers. *Developmental Science*, 7, F1-F9.
- Mueller, V. A., Brass, M., Waszak, F., & Prinz, W. (2007). The role of the preSMA and the rostral cingulate zone in internally selected actions. *NeuroImage*, 37, 1354-1361.
- Obhi, S. S., & Haggard, P. (2004). Internally generated and externally triggered actions are physically distinct and independently controlled. *Experimental Brain Research*, 156, 518-523.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 255-258.
- Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica*, 124, 106-128.
- Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the intentions of others: The perceived

- intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, 16, 1706-1716.
- Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain*, 128, 1038-1048.
- Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia*, 41, 156–70.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience*, 1, 245-258.
- Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, 308, 214-216.
- Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness and autism. *Child Development*, 76, 502-517.
- Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, 36, 1-16.
- Premack, D., & Premack, A. J. (1995). Origins of human social competence. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 205-218). Cambridge, MA: MIT Press.
- Premack, D., & Premack, A. J. (1997). Infants attribute value +/- to the goal-directed actions of self-propelled objects. *Journal of Cognitive Neuroscience*, 9, 848-856.
- Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S. J., & Passingham, R. E. (2000). The prefrontal cortex: Response selection or maintenance within working memory? *Science*, 288, 1656-1660.
- Sabbagh, M. A., Bowman, L. C., Evraire, L. E., & Ito, J. M. B. (in press). Neurodevelopmental correlates of theory of mind in preschool children. *Child Development*.
- Santos, L. R., Morticorena, D. C. W., & Goddu, A. (2007). Do monkeys reason about the false beliefs of others? Paper presented at the biennial meeting of the Society for Research in Child Development,

Boston, MA.

- Santos, L. R., Nissen, A. G., & Ferrugia, J. A. (2006). Rhesus monkeys, *Macaca mulatta*, know what others can and cannot hear. *Animal Behavior*, *71*, 1175-1181.
- Saxe, R., Carey, S., & Kanwisher, N. (2004a). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, *55*, 87-124.
- Saxe, R., & Kanwisher, N. (2003). People thinking about people: The role of the temporo-parietal junction in 'theory of mind.' *NeuroImage*, *19*, 1835-1842.
- Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, *43*, 1391-1399.
- Saxe, R., Xiao, D-K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004b). A region of right superior temporal sulcus responds to observed intentional actions. *Neuropsychologica*, *42*, 1435-1446.
- Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. *NeuroImage*, *35*, 1378-1384.
- Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*, B1-B11.
- Song, H., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology*, *44*, 1789-1795.
- Song, H., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an actor's false belief be corrected by an appropriate communication? Psychological reasoning in 18.5-month-old infants. *Cognition*, *109*, 295-315.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by two-year-olds. *Psychological Science*, *18*, 587-592.
- Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs to 13-month-old infants. *Psychological Science*, *18*, 580-586.

- Tager-Flusberg, H. (2005). What neurodevelopmental disorders can reveal about cognitive architecture: the example of theory of mind. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The Innate Mind: Structure and Contents* (pp. 272-288). New York: Oxford University Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675-735.
- Tomasello, M., & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what's new for other persons. *Developmental Psychology*, 39, 906-912.
- Waszak, F., Wascher, E., Keller, P. E., Koch, I., Aschersleben, G., Rosenbaum, D. A., et al. (2005). Intention-based and stimulus-based mechanisms in action selection. *Experimental Brain Research*, 162, 346-356.
- Wellman, H. M. (in press). Developing a theory of mind. To appear in U. Goswami (Ed.), *The Blackwell handbook of cognitive development (2nd Edition)*. Oxford: Blackwell.
- 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.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1-34.
- Woodward, A. L. (2005). The infant origins of intentional understanding. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 33, pp. 229-262). Amsterdam: Elsevier.
- Wu, D., & Baillargeon, R. (2007). Can an agent be inert? 14-month-old infants' reasoning about agency without motion cues. Paper presented at the biennial meeting of the Society for Research in Child Development, Boston, MA.

Figure Captions

Figure 1: Schematic drawing of the events shown in the first two and last two familiarization trials in the left-right condition of Experiment 1.

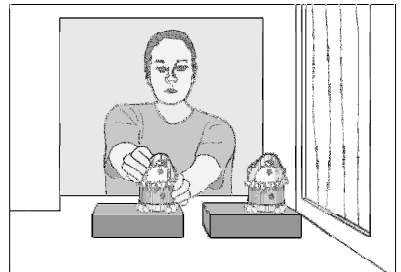
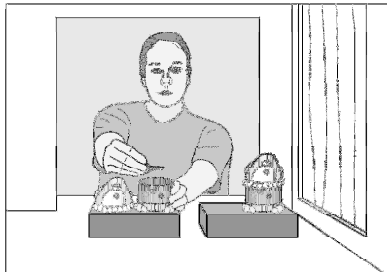
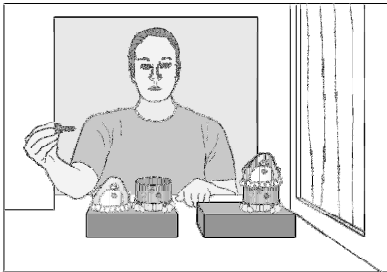
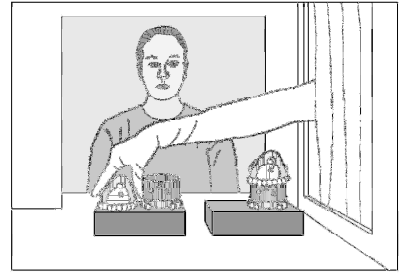
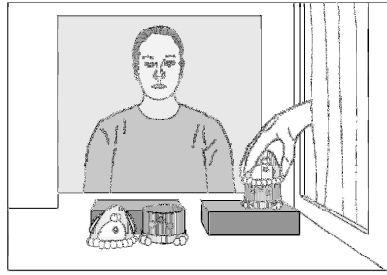
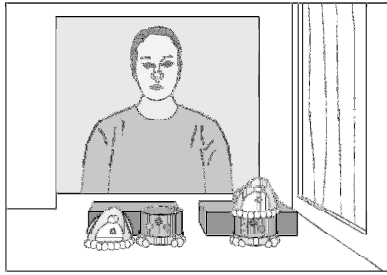
Figure 2: Schematic drawing of the transparent- and opaque-cover test events in the false-belief condition of Experiment 1.

Figure 3: Schematic drawing of the transparent- and opaque-cover test events in the true-belief condition of Experiment 1.

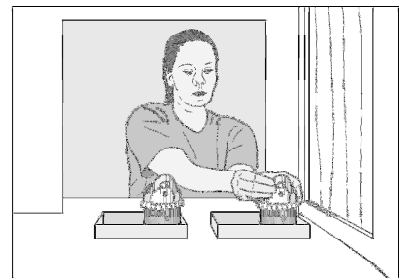
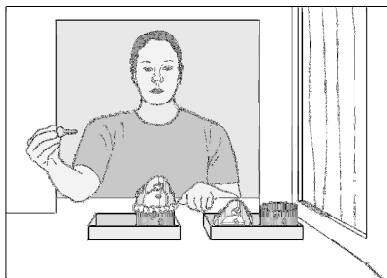
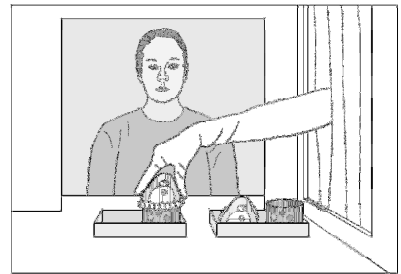
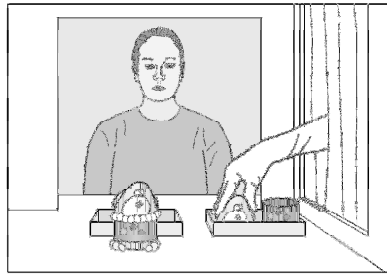
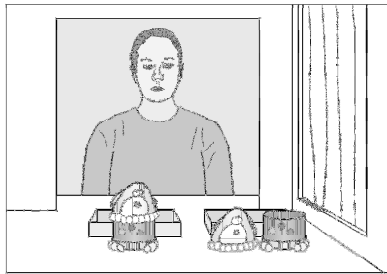
Figure 4: Mean looking times to the test events in the false- and true-belief conditions of Experiment 1, in the false- and no-key conditions of Experiment 2, and in the ignorance condition of Experiment 3. Error bars represent standard errors.

Figure 5: Schematic drawing of the final phases of the correct- and incorrect-cover test events in the transparent- and opaque-covers ignorance conditions of Experiment 3.

Familiarization Events 1 and 2



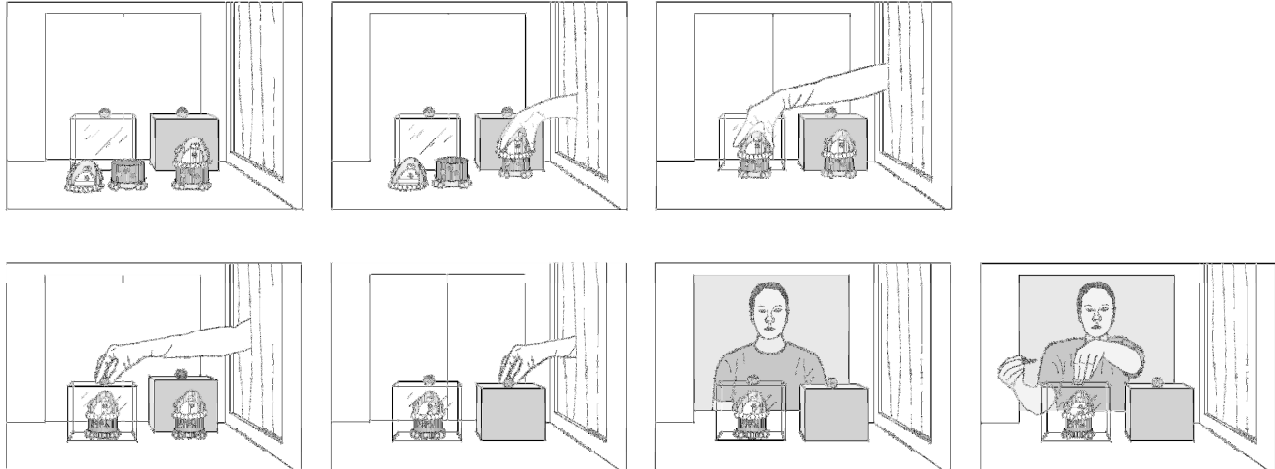
Familiarization Events 3 and 4



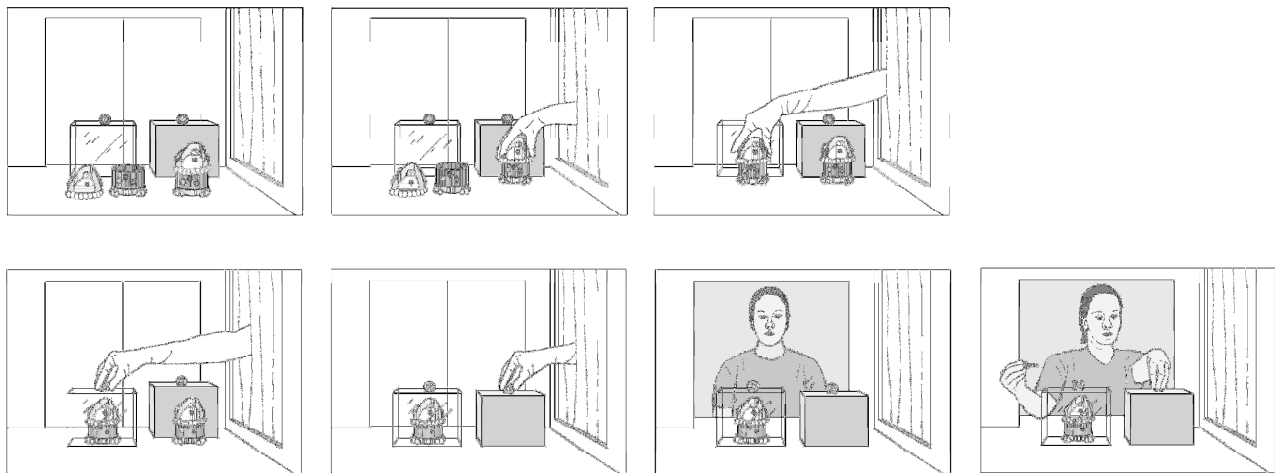
False-belief Condition

Test Events

Transparent-cover Event



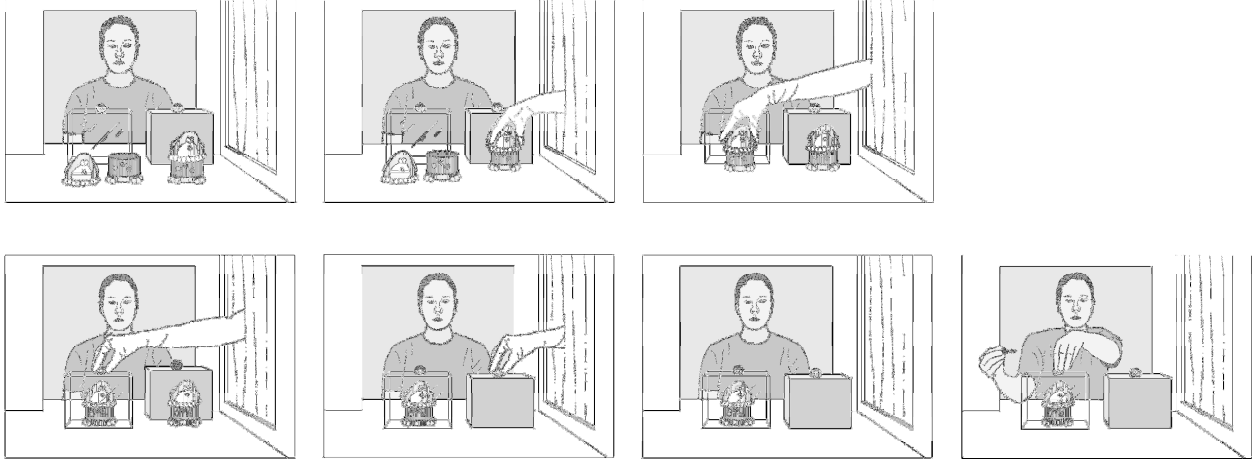
Opaque-cover Event



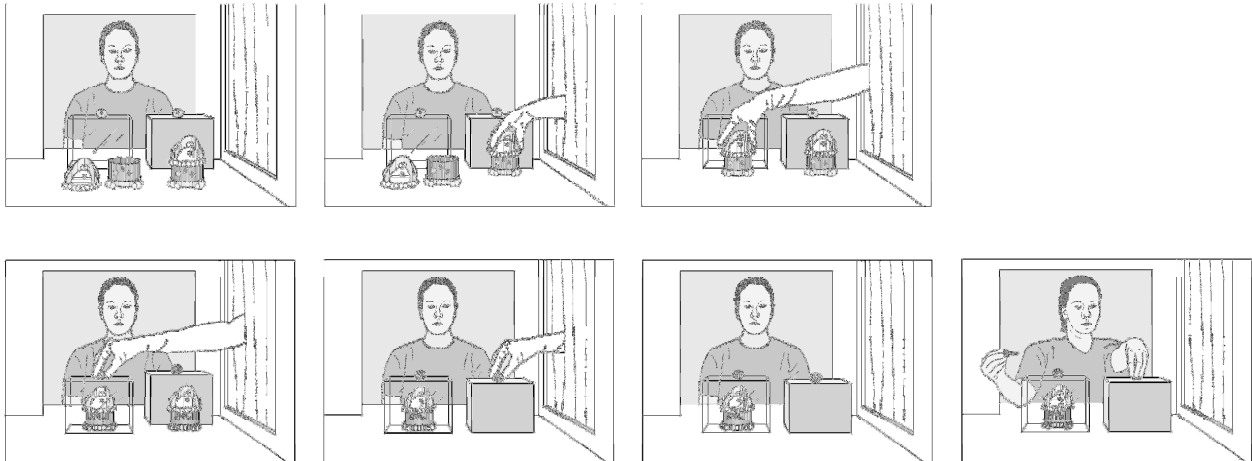
True-belief Condition

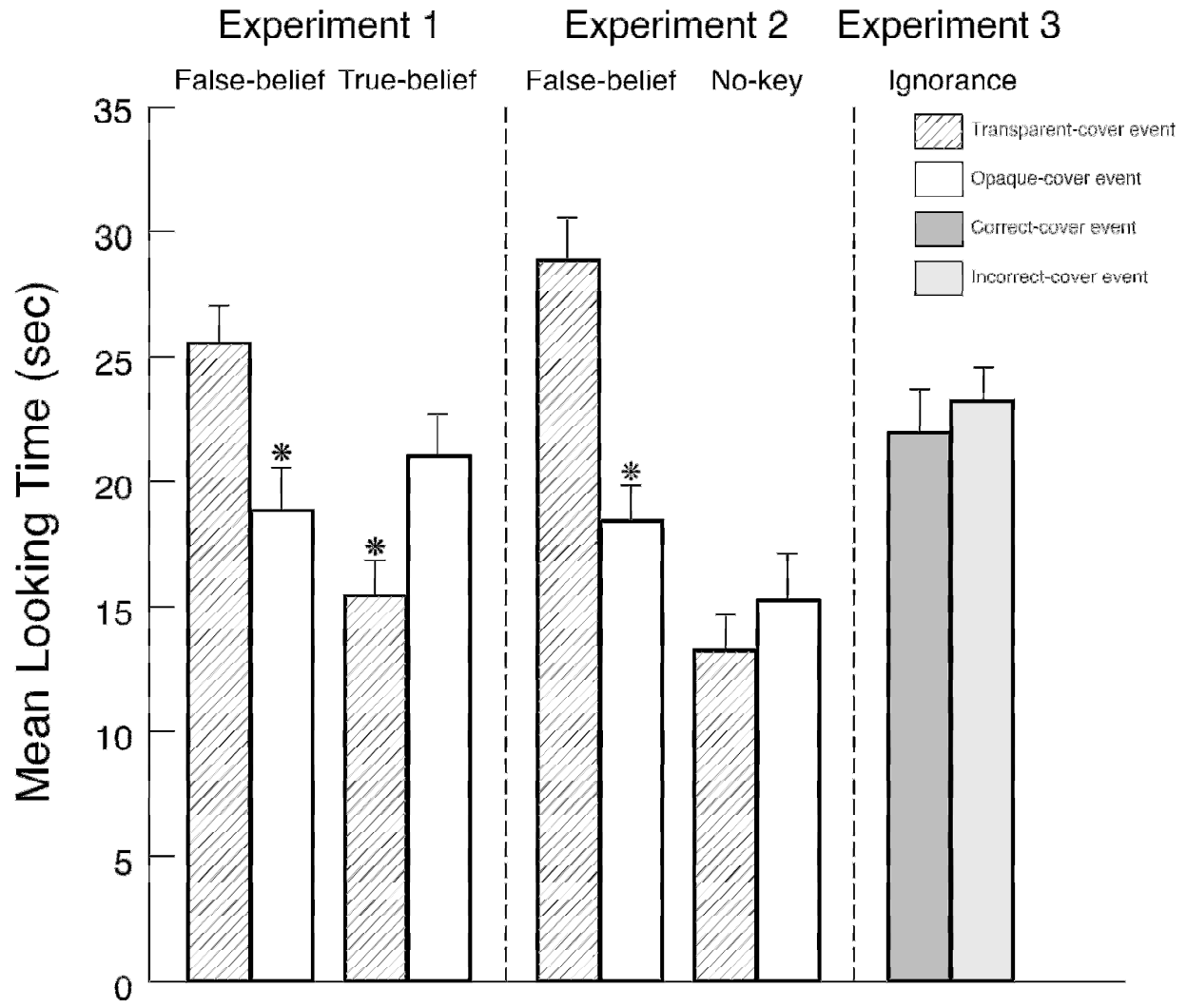
Test Events

Transparent-cover Event



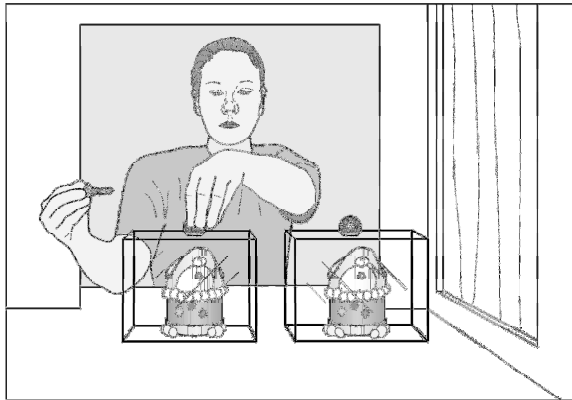
Opaque-cover Event



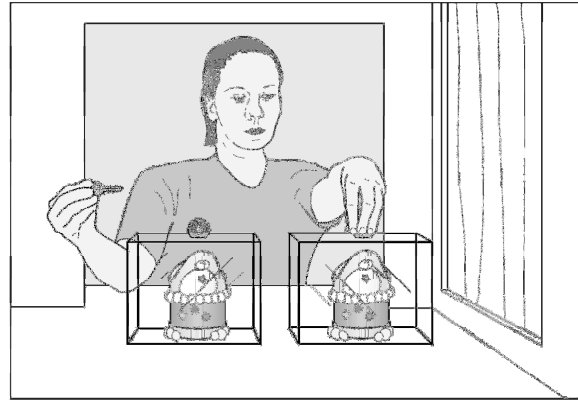


Ignorance Condition

Transparent-covers Condition

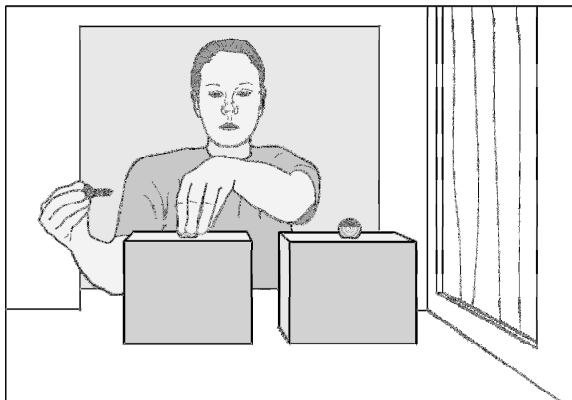


Correct-cover event

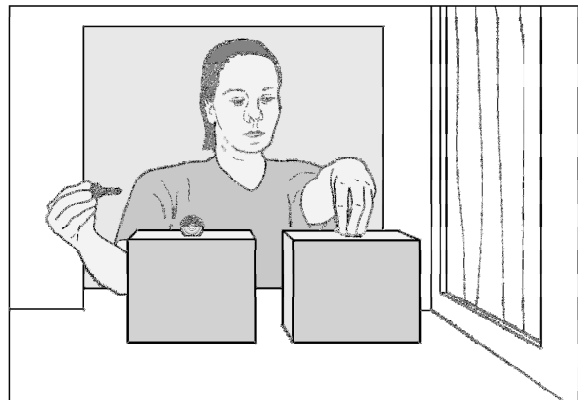


Incorrect-cover event

Opaque-covers Condition



Correct-cover event



Incorrect-cover event