

Understanding Other Minds: Perspectives from developmental social neuroscience

Simon Baron-Cohen, Michael Lombardo, and Helen Tager-Flusberg

Print publication date: 2013

Print ISBN-13: 9780199692972

Published to Oxford Scholarship Online: September 2013

DOI: 10.1093/acprof:oso/9780199692972.001.0001

Early manifestations of mindreading

Victoria Southgate

DOI:10.1093/acprof:oso/9780199692972.003.0001

Abstract and Keywords

Recent research has forced a rethink of the commonly held consensus that young children have little understanding of other minds before the age of four years. With novel methods and new paradigms, it is now clear that even young infants are sensitive to others' perspectives and resulting beliefs. What does such a sensitivity imply? This chapter reviews the evidence to date and interpretations of the data that have emerged. It then asks how we might characterize an early sensitivity to beliefs by looking at what behaviours these early representations might support. Finally, it points out ways in which infant's understanding of other minds might permit inferences about behaviour without including some of the attributes of a concept of belief that, as adults, we readily appreciate.

Keywords: Infants, early theory of mind, false belief, looking-time, action prediction

A decade ago, there was no chapter within a book on understanding other minds devoted to the infancy period, because it was generally agreed that no such understanding existed at this stage of development. Traditional tests of mindreading, like the Sally-Anne or Smarties task (Gopnik & Astington, 1988; Baron-Cohen, Leslie & Frith, 1985; Perner, Leekam & Wimmer, 1987; Wimmer & Perner, 1983), were robustly failed by children under the age of 4, and it was concluded that, until this point, children essentially lacked an appreciation that other's behavior can be driven by unobservable mental states (Wellman, Cross & Watson, 2001). While some authors maintained that younger children's failure on these classic tasks likely reflected performance issues, rather than any conceptual deficit (Leslie, 1994, Leslie & Polizzi, 1998), modifications aimed at

enabling young children to overcome performance limitations did not lower the age at which children passed these tasks to any great degree (e.g. German & Leslie, 2000). Nevertheless, it was puzzling that, despite failing these tasks, young children's behavior suggested a greater understanding of other minds than their explicit task performance would give them credit for. For example, one influential theory of communication holds that it involves the ability to represent others' mental states (Grice, 1989). This is clear when we consider the case of pronouns: using terms like "it", or "he" suggests that the person using those terms considers that the other person shares their understanding of what "it" is, or who "he" is, and these pronouns are abundant in the speech of toddlers (Bloom, 2000).

It has long been known that the method used to test for the presence of an understanding has an influence on whether or not such understanding is revealed. For example, relying on infants explicit search behavior would lead us to believe that they have no appreciation of object permanence before 7 months, when they begin to uncover hidden objects themselves (Piaget, 1954). However, researchers using looking-time (the length of time that infants look toward outcomes that are expected or unexpected if one possesses a particular concept), and anticipatory looking have demonstrated that, in fact, infants as young as 3 months grasp object permanence (Baillargeon, Spelke & Wasserman, 1985; Ruffman, Slade & Redman, 2005). In a move paralleling that previously seen in the domain of physical cognition, the last decade has seen researchers turn to infants looking behavior, as a measure of their social cognitive abilities. In findings paralleling those in the domain of physical cognition, infants looking behavior reveals a far greater appreciation that other's behavior is generated by their own representations of the world, than classic tests of mind-reading would have us believe.

Evidence for mindreading in infancy

The first forays in to using infants looking behavior as an indication of what they understand about other minds investigated whether infants understand that others' movements are directed toward their goals (e.g. Gergely, Nadasdy, Csibra, & Biro, 1995; Woodward, 1998). In one of the most well replicated results in the domain of infant social cognition, Woodward (1998, 1999) **(p.4)** demonstrated that 6- and 9-month-old infants encoded a relationship between an agent and the target of its actions, and reacted with longer looking when the agent subsequently acted on a new target object. While these findings demonstrate that infants encode actions in terms of what they are directed toward (their intentionality; Gomez, 2008), and expect that having acted on a target multiple times, people will most likely continue to act on the same object again, they do not tell us whether infants are appealing to unobservable mental states to generate these expectations. There is no need to attribute to infants an understanding that others' actions are generated by unobservable mental states like "goal", as alternative non-mentalistic interpretations are equally plausible

(Gergely & Csibra, 2003; Ruffman, Taumoepeau & Perkins, 2011). Thus, evidence that young children are using others' mental states in order to generate expectations about their behavior needs to come from situations in which the child and the other hold a different representation of the world (Dennett, 1978), such as is the case in the classic false-belief task.

False belief understanding in infants

In the first study to suggest that infants are capable of representing other's unobservable mental states, Onishi & Baillargeon (2005) cleverly transformed the classic Sally-Anne task (Baron-Cohen et al., 1985; Wimmer & Perner, 1983) into a looking-time paradigm. Infants observed as an actor placed an object in to one of two boxes and then reached in to retrieve her object. After that, infants saw the agent disappear, during which time the object moved, by itself, to the opposite box. Infants then observed one of two outcomes. In the congruent test trial, the actor then reached in to the box where she had left her object (congruent because she could not have known that the object had moved in her absence and so she should reach in to the box where she left it), but in the incongruent test trial, she reached in to the box where the object actually is (incongruent because, unaware that the object had moved, she should not reach to its real location). Onishi & Baillargeon (2005) found that infants look significantly longer when they watch the incongruent outcome than the congruent outcome, suggesting that they find someone searching in the location where the object really is (when they could not know where it really is) unexpected—and the fact that they respond with longer looking times toward an outcome in which the person searches where the object actually is, suggests that infants have some understanding that people's actions should depend on what they have experienced, rather than on what is actually the case.

One of the objections to the conclusions drawn by Onishi & Baillargeon (2005), that their finding demonstrates that infants understand that others act on the basis of their representations, was that any understanding of mental states should manifest in a variety of contexts, and that we are only permitted to conclude that 4-year-olds operate with a mentalistic understanding of others because such understanding has been demonstrated in many different contexts and scenarios (Perner & Ruffman, 2005). In answer to this challenge, many studies have now appeared which do extend this original finding to different contexts. For example, Song and Baillargeon (2008) showed that infants understand that others can have representations concerning not only the location of an object, but also its contents. In a non-verbal task reminiscent of the classic "Smarties" task (Perner et al., 1987), they showed infants that an actor preferred to reach for a doll with blue hair than a skunk. Then, when the actor was absent, someone put the doll in a plain box and the skunk in a box that had blue hair sticking out of it. Like children who pass the Smarties test, 14.5-month-old infants seemed to expect the actor to be misled by the visible tuft of hair and looked longer when she searched in the plain box, even though that was

actually where her desired **(p.5)** doll was located. There are now numerous other examples of different contexts in which infants appear able to think about what other people have experienced (e.g. Luo, 2011; Song, Onishi, Baillargeon & Fisher, 2008; Scott, Baillargeon, Song & Leslie, 2010; Surian, Caldi, & Sperber, 2007; Yott & Poulin-Dubois, 2012).

What kind of behavior do these representations support?

Looking-time studies have demonstrated that infants are sensitive to the fact that others' actions are motivated by their own view of the world, even when this is different from what the infant should know to be the real state of the world. In other words, preverbal infants appear to represent another person's representation of the world. What kind of behavior might this ability to represent other's perspectives on the world support? It is often held that one of the primary reasons why we may need to engage in mindreading is to enable us to generate predictions concerning what other people may do (Premack & Woodruff, 1978; Dennett, 1978), and the ability to accurately predict others' actions is crucial for any social species (Verfaillie & Daems, 2002). However, because measures like looking-time rely on infant's responses to the outcomes of events, they cannot tell us whether infants are able to use this sensitivity to others' mental states, to generate accurate predictions about their behavior. While looking time is assumed to rest on violated expectations (or predictions) that are formed in advance of the infant seeing an outcome, it may equally reflect the infant's recognition that an outcome is incongruent with a preceding event only when the outcome is seen (Southgate & Csibra, 2009). Thus, while looking time certainly tells us that infants can recognize behavior that is inconsistent with a perspective that an agent should hold, it does not tell us whether infants can use their sensitivity to others' perspectives usefully, to generate predictions about what they may consequently do.

In a first step toward addressing this question, we used eye-tracking to ask whether 25-month-old infants could predict where an agent will search for a desired object when she has a false belief about that object's location (Southgate, Senju & Csibra, 2007). In this task, infants first saw two familiarization trials in which an agent, sitting behind a panel with two windows and two boxes (Figure 1.1), observed as a puppet placed an object in to one of the two boxes. The puppet then disappeared, and both windows lit up at the same time as a chime was heard—at which point the agent reached through the window located behind the box containing the toy, and opened the box. These two familiarization trials served to show the infant that this cue (both windows lighting up and chime) signaled the imminent reach of the agent through one of the windows. In the test trial, infants then saw a false-belief scenario in which, once the puppet had placed the toy in one of the boxes, the agent turned around, and did not witness the puppet subsequently remove the toy from the box and take it away. The question was, when the agent turned back toward the boxes, could infants use their apparent sensitivity to others' perspectives on events to

predict that the agent will reach through the window above the box where she saw the toy being placed, even though the infant has seen that the toy is no longer there? Results showed that, indeed, they could. When the agent turned around after the toy was placed in the left-hand box, infants predicted she would open the left-hand window upon her return, and when she turned around after it was placed in the right-hand box, they expected her to open the right-hand window.

More recently, Knudsen & Liszkowski (2012a) have also demonstrated that even younger infants, at 18 months, generate predictions in accord with others' epistemic states. In their study, they use what they call an "anticipatory intervening" paradigm, originally suggested by Dennett (1978), in which infants observed as an actor re-entered a room with either a true- or false-belief about the location of her toy. When the actor had a false-belief that her toy was in the box where she **(p.6)**

(p.7) left it, 18-month-olds intervened to warn the actor of the real location of the toy, before she had approached the location where she thought the toy was, but they did not do this when she had a true belief. Together, these studies suggest that, by 18 months, infants can use their understanding that others' actions are modulated by their epistemic states, to generate predictions about what actions will follow, and even prepare an appropriate response (e.g. pointing to inform the other) for that prediction. Mindreading or behavior reading?

The data reviewed so far have been interpreted as reflecting infant's appreciation that unobservable mental states modulate others' actions (e.g. Baillargeon, Scott, & He, 2010). However, a constant challenge for those advocating that infants are encoding others' mental states, is that simpler, non-mentalistic interpretations in terms of behavior-reading are always possible (Low & Wang,

2011; Ruffman et al., 2011). The problem is that one cannot know for sure whether infants are genuinely using mental state concepts such as "knows" or "believes" when generating predictions about what the other will do, or whether they have simply learnt associations between events and typical behaviors that follow from those events (Ruffman et al., 2011). Since mental states drive behavior, every mental state will have a behavioral correlate. One commonly suggested probability that children might exploit is that people who are ignorant of something will usually get the wrong answer (e.g. Perner & Ruffman, 2005). If children evoke such an "ignorance rule" in the false-belief task (e.g. the Sally-Anne task), they would reason that, because Sally did not see Anne move the toy from location A to location B, Sally will probably get the wrong answer (i.e. she will choose location A because the infant knows that the object is actually in location B). However, crucially, this does not entail that the child attributes to Sally a mental representation that the toy is at location A—she will simply search there because this is not where the object is. While the evidence that children

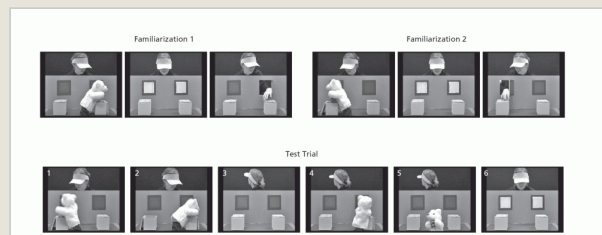


Figure 1.1 Upper panel: The two familiarization trials in which the puppet places a ball in either the right or left box. After the puppet leaves the scene, the two windows are illuminated and shortly after the actor reaches through the window above the box in which the puppet has left the ball. Lower panel: Test trial. The puppet places the ball in the left-hand box (1) and then removes it and places it in the right-hand box (2). The actor then turns around (3) and the puppet returns to remove the ball from the right-hand box (4 and 5). The actor then reorients to the scene and the windows are illuminated. Infants looking behaviour is measured from the onset of the illumination. A second group of infants received an alternative test trial in which the correct location for the actor to search was the left-hand box.

evoke such a rule is somewhat contradictory (e.g. Friedman & Petrashek, 2009; Hedger & Fabricius, 2011; Ruffman, 1996), it does not seem that infants are using this rule in solving these tasks. In our previously described study (Southgate et al. 2007), infants saw a scenario in which the object was actually removed from the scene altogether by the puppet (rather than being transferred from A to B), rendering both boxes the wrong place to search if you actually want to find the ball. Infants do not expect the agent to search indiscriminately, however; they specifically expect her to search in the location where she last saw the object, suggesting that they attribute to her an epistemic state in which the object is in location A. Similarly, in a follow-up to their original study (Knudsen & Liszkowski, 2012a), Knudsen & Liszkowski (2012b) again used their “anticipatory intervening” paradigm to tease apart responses based on an ignorance rule and those based on an attributed epistemic state. As in Southgate et al., (2007), the agent’s toy was removed entirely during the period when the agent was absent and, instead, aversive objects that the infant knew the agent did not like, were placed in each container. When the agent returned, infants pointed more to the location where the agent had left her toy, suggesting that they expected her to specifically choose one location. If infants were operating with an ignorance rule, since both locations were wrong (as the object had been removed, rather than transferred), they should have pointed to both locations. In a final condition, the agent was, in fact, ignorant of the location of her toy and, in this case, infants pointed equally to both locations, suggesting that they understood that her ignorance should lead her to search indiscriminately. Recently, looking-time studies have also begun to include conditions that control for the use of such an ignorance rule, and confirm that infants have different expectations depending on whether the agent should have a specific (false) belief about the location of the object, or is simply ignorant of its location (e.g. Scott & Baillargeon, 2009).

(p.8) While infants do not seem to be relying on an ignorance rule, nevertheless, there are other behavioral rules or probabilities that infants might have acquired. For example, in making their case against mindreading in non-human primates, Povinelli & Vonk (2003) argue that, while these animals might seem like they are sensitive to others’ mental states, they may simply have learned that people tend to search for things in locations toward which they were last oriented (see Perner & Ruffman, 2005 for the suggestion that infants may exploit the same rule). For Povinelli & Vonk (2003), learning that orientation is correlated with subsequent search is sufficient to form the expectation that people will search for things where they last saw them. So, if an actor is oriented toward the left-hand box when the object was placed in it, and this was the last point at which her eyes were oriented toward the object, this is where she will probably search for it. Since we usually do look for things where we last put them, or saw them, it is plausible that infants might acquire such a rule through observing the behavior of others.

Of course, this kind of rule- or probability-based explanation applies equally to the standard false-belief task which is taken as evidence that older children operate with a representational theory of mind, and are not just good behavior readers. However, as older children can, in addition to generating correct predictions, also explain why Sally will search in the false-belief location, and give verbal explanations consisting of mental state terms like “thinks” and “knows,” we can be more confident that they are not solving the tasks merely by employing these kind of behavioral rules. How could we know whether infants are really considering the unobservable mental states of others in generating predictions about their behavior, or whether these predictions are based on associations between events and behavior?

One strategy has been to argue that, because infants succeed on a whole host of different tasks, and each would require a different learned association to be solved by applying different behavioral rules, it is simply more parsimonious to assume that they are solving these tasks by appealing to common mental states (Baillargeon et al., 2010).¹ Another strategy has been to directly test whether young children inflexibly use such rules. For example, Trauble, Marinovic, & Pauen (2010) directly tested the hypothesis that infants solve non-verbal false-belief tasks by using the “people search in the last place that they saw something” rule. To address this, Trauble and colleagues (2010) ran a study that was largely similar to Onishi and Baillargeon (2005), with the exception that, in one condition (the manual control condition) the agent gained their knowledge through touch, rather than vision. While the agent was turned around, she manipulated a balance beam behind her back, which caused the object to roll in to the opposite box. Thus, even though she did not actually see the object move (and so the last place she saw the object was not the location where it now is), her own manipulation caused the object to move, and so she should know its actual location. As Trauble and colleagues (2010) argue, if infants rely blindly on a rule like “people search for things where they last saw them,” the agent should still be expected to look for her object in the location where she last saw it, even though she should know that it is in the opposite box since it was her who caused its movement. Infants are not fooled by this: they do not expect her to search for the ball in the location where she last saw it, instead they look longer when she searches in this location than in the location where the object actually is (just as infants in a true-belief condition did), presumably because they understand that the agent should know where the ball actually is.

Nevertheless, Povinelli & Vonk (2003) are right when they argue that “no experiment in which theory of mind coding derives from a behavioral abstraction will ever suffice” to definitely answer **(p.9)** the question of whether non-verbal subjects are utilizing behavioral rules, or appealing to other’s mental states. Even if infants in Trauble and his colleagues’ (2010) study were not relying on a “seeing” rule, they may have been relying on some other rule involving “touching” (Low & Wang, 2011). Povinelli and Vonk’s point is that, because every

mental state will have a behavioral correlate, any study in which mindreading skills are assessed by comparing performance on conditions which vary in a behavior (e.g. true belief vs. false belief), can always be explained in terms of rules. Since we simply do not know how much experience infants need in order to form behavioral correlations, appealing to parsimony to solve this dilemma is unsatisfying. However, Povinelli & Vonk (2003) do advocate a possible solution following an earlier suggestion by Heyes (1998). Specifically, in order to tease apart genuine appreciation of unobservable mental states from reliance on behavioral associations, one needs to create a situation in which expectations are based on an experience for which it would not have been possible for the subject to have acquired any behavioral correlate. This could be achieved by providing the subject with direct first-person experience of one of its own mental states—seeing—and asking whether they can extrapolate this subjective experience to another individual, even though they would have no opportunity to acquire any behavioral correlate of this mental state. A version of this paradigm was recently implemented by Meltzoff and Brookes (2008) in which they provided 12- and 18-month-old infants with differential experience with a blindfold. For one group of infants, the blindfold was opaque and prevented seeing; for the other group, a trick blindfold which, although perceptually identical to the opaque blindfold, did allow the infant to see. They tested infants' abilities to extrapolate their first-person experience with the blindfold to others, by asking whether, when infants saw another person wearing the blindfold (all infants actually saw the experimenter wear the opaque blindfold), they would selectively follow the “gaze” direction of that person, depending on which blindfold they had experienced. Their results showed that infants who had experienced the trick blindfold subsequently followed the gaze of the experimenter who was wearing the blindfold significantly more than infants who had experienced the opaque blindfold, suggesting that, indeed, infants had extrapolated their own first-person experience of seeing, or not seeing, to another person.

In a recent study, we extended this methodology to a false-belief scenario with 18-month-olds, and asked whether differential blindfold experience not only modulated what infants understood about what the other could see, but also their predictions about what the other would do (Senju, Southgate, Snape, Leonard & Csibra, 2011). As in Meltzoff and Brooks (2008), we provided two groups of infants with either an opaque blindfold experience, or a trick (transparent) blindfold experience. Following the blindfold experience, all infants took part in an anticipatory looking paradigm similar to that used by Southgate et al. (2007). The main difference between this study and the previous study is that, while in Southgate et al. (2007) infants saw the agent turn around while the puppet removed the object from the box, in this version the agent donned the blindfold.² We hypothesized that if infants understood from their own experience that the blindfold was opaque or transparent, then this should

modulate whether they infer that the agent has seen the puppet remove the object or not and, consequently, whether the agent will search in the location where the object was before she put on the blindfold. Specifically, we hypothesized that infants who had the opaque blindfold experience would understand that the agent could not see the puppet removing the object, and so should expect her to search in the location where she last **(p.10)** saw the object (the false-belief location), whereas we hypothesized that infants who had the trick blindfold experience would understand that the agent could see the puppet removing the ball and so would have no particular expectation about where she would search. This is exactly what we found. Fourteen out of 18 infants who had undergone the opaque blindfold experience expected the agent to search in the false-belief location (a result significantly above chance), whereas only six out of 18 infants who had the trick blindfold experience looked in anticipation toward the false-belief location. These results show that, not only did infants use self-experience to assess what the other person could see, they understood the causal role that seeing has in generating action. Crucially, these results cannot be explained in terms of a reliance on behavioral rules because, since infants never saw the adult wearing the blindfold, they had no opportunity to acquire any behavioral correlate of blindfold wearing.³

Beyond behavioral rules

The evidence reviewed above suggests that infants are not solving these false-belief scenarios by a reliance on learned behavioral rules and that they are, indeed, representing the world from the perspective of the other. However, to what extent does infants' success on these tasks reflect an appreciation of mental state concepts as they are represented by adults, and what might infants' representations look like? One of the biggest challenges is to account for why, if infants do possess a genuine representational theory of mind, do children younger than 4 years not pass the standard false-belief tasks?

Several theorists have attempted to explain this paradox. One theory is that young infants have a representational theory of mind, with concepts like seeing, knowing and believing, but are unable to demonstrate this knowledge on the standard false-belief tasks (German & Leslie, 2000; Baillargeon et al., 2010). For example, Baillargeon and colleagues argue that in order to respond correctly on the verbal false-belief task, infants would need not only to represent the other's false belief, but additionally they would need to access this representation in order to select the correct response, and simultaneously inhibit a prepotent tendency to answer the question based on what they know to be true (Baillargeon et al., 2010; Scott & Baillargeon, 2009). This position advocates that infant's representations of other minds are not different from adults, but due to limited resources, or immature neural connections, they cannot do all three things (represent beliefs, select the right answer and inhibit the wrong answer) at once.

There are several problems with this account. Fundamentally, this account is open to similar criticisms as previous accounts that involved explaining infants' failures by appealing to limited executive resources. By designing tasks that alleviate some of the executive demands, several researchers have succeeded in lowering the age of passing false-belief tasks by a few months, but there is no dramatic change (e.g. Carlson, Moses & Hix, 1998; Surian & Leslie, 1999). A bigger problem is that it is not clear why performance on the so-called "spontaneous-response tasks" should not also require these additional elements that Baillargeon and colleagues (2010) reserve for the elicited response tasks. For Baillargeon and colleagues (2010), infant's performance on their non-verbal false-belief tasks suggest that infants "realize that others act on the basis of their beliefs and that these beliefs are representations that may or may not mirror reality" (Onishi & Baillargeon, 2005). If it is, indeed, the case that infants performance on non-verbal false-belief tasks reflects this kind of understanding **(p.11)** (including an appreciation that the agent's belief is a representation that is false), then presumably some response inhibition would need to be deployed in order for the infant not to respond based on what the infant knows to be the true location of the object? In Baillargeon and colleagues' (2010) response account, it is not clear why a saccadic response (e.g. Senju et al., 2011; Southgate et al., 2007) should be immune to an egocentric bias, but a verbal response should not be.

A second problem for this account is that more recent paradigms, in which infants also appear to consider others false beliefs, would appear to involve the elements of response selection and response inhibition that Baillargeon and colleagues (2010) argue are too challenging for infants limited resources. In one task, Southgate, Chevallier and Csibra (2010) presented 17-month-old infants with a paradigm in which, in order to select the object about which the experimenter was referring, infants needed to consider where the experimenter believed the object to be. Infants saw the experimenter placing a different novel object, one in each of two similar boxes, and closed the lids. The experimenter then left the room, after which a new experimenter suddenly appeared from behind the curtains at the back of the room, crept over to the boxes and switched their contents (i.e. the object in the left-hand box was transferred to the right-hand box, and vice versa). The new experimenter then closed the lids again and crept back behind the curtains. At this point, the original experimenter returned to the room, knelt down behind the two boxes and pointed toward one of the closed boxes and said to the infant "Do you remember what I put in here? Can you get it for me?".⁴ Since the experimenter was not in the room when the objects were switched, infants should reason that she is *not* now referring to the object that is actually in the box toward which she is pointing, but the object that was in the box and is now in the other box. So, the correct response is to give the experimenter the object in the non-referred-to box. The results showed that this is what infants did: nine out of 12 infants

selected the object in the non-referred box in response to the experimenter's request. On the other hand, in a true-belief control condition in which the first experimenter stayed in the room and witnessed the second experimenter switching the contents of the boxes, nine out of 12 infants chose the box that the experimenter was actually pointing to. In this case, because she saw the switch and knew the contents of the boxes, infants took her pointing as referring to the object that was now in the box.

According to Baillargeon and colleagues, this kind of "indirect elicited-response" task would not demand the same kind of response selection and inhibition processes required by the standard false-belief task because there is no question that directly taps their representation of the agent's belief (Baillargeon et al., 2010). It is not clear on what grounds Baillargeon and colleagues (2010) draw the conclusion that a direct (e.g. standard Sally-Anne task) and an indirect (e.g. Southgate et al., 2010) elicited-response task would require differences in response-selection and response-inhibition. Both tasks require an explicit⁵ response from the infant, and in order to generate their response, infants need to consider what the agent thinks about the location of the objects. Moreover, it could be argued that the Southgate et al. (2010) task is also demanding of inhibitory processes. In this task, infants need to ignore the location toward which the agent is pointing, and thus made most salient, and select the other box that the experimenter is ignoring. **(p.12)** Thus, presumably, some degree of response-inhibition is required for this task, and there seems no a priori reason to assume that, if response inhibition is required to pass the standard Sally-Anne task, it is greater than would be required to pass this kind of "indirect" elicited-response task (see also Buttelmann Carpenter, & Tomasello, 2009).⁶

Thinking about what is involved in passing false-belief tasks

Evidence reviewed in previous sections should give us reason to doubt that the gap between infants passing non-verbal theory of mind tasks, and the age at which children pass standard theory of mind tasks can be explained either by appealing to behavioral rules (Perner & Ruffman, 2005), or the elimination of response selection and inhibition (Baillargeon et al., 2010). In the absence of a convincing explanation for why infants pass non-verbal false-belief tasks, but older children still fail verbal false-belief tasks, we should think carefully about just what is involved in success on these different tasks.

According to Scott, Baillargeon and colleagues' response account, infants possess a decoupling mechanism that enables them "to hold in mind two distinct versions of a scene: one that corresponds to reality, and one that incorporates the agent's false belief" (Scott et al., 2009). Furthermore, they argue that infants realize that other's beliefs may not mirror reality (Onishi & Baillargeon, 2005). While they argue that response inhibition is not required for non-verbal false-belief task success, others disagree and argue that the lack of a bias toward responding based on the reality-based representation is puzzling given

that (a) infants have notoriously poor inhibitory capacities and (b) even adults default to a reality-based interpretation of others' behavior under certain circumstances (Ruffman & Perner, 2005; Samson & Apperly, 2010).

While it is often assumed that success on false-belief tasks (verbal or non-verbal) requires that children realize that others' beliefs are distinct from their own (e.g. Carpendale & Lewis, 2006; Onishi & Baillargeon, 2005), this is not necessarily so. In early discussions of what constitutes an understanding of other minds, the emphasis was on representing other's representations (Dennett, 1978) without emphasis of the need to appreciate the fact that other's representations might conflict with reality. The development of the false-belief task evolved not out of a need to demonstrate any appreciation that beliefs can be false, but rather because predictions based on beliefs that are true, while they might reflect mental state reasoning, might just as easily reflect infant's expectation that others will behave in accord with reality. To borrow Dennett's Punch and Judy example, the fact that 4-year-olds squeal in anticipatory delight when they see Punch about to push the box (that he thinks contains Judy, even though the children have seen her escape) off the cliff suggests they were able to reason that he is acting on a mistaken belief that Judy is still in the box. They laugh because they know that Judy is not in the box, but that Punch thinks she is. It is only funny because children know both of these things: they know that Punch has a belief, and *that* his belief is false.

Do infants need to appreciate both of these things in order to generate a prediction concerning where an agent will search when she hasn't seen the displacement of her desired toy? The answer is no, not necessarily. In order to generate a correct prediction about where someone will search for an object, infants need only to consider the agent's representation of the object's location. They do not need to understand or appreciate that her representation is false, or think about how it compares with reality. To illustrate, in Southgate et al. (2007), infants observe an agent turn around after a puppet has put a toy in the left-hand container. During the time she is oriented away from the scene, the **(p.13)** puppet returns to take the toy away. To correctly anticipate that the agent will search in the left-hand box upon her return, infants might reason that she will search in the left-hand box because she last saw it there even though it is actually now in the right-hand box. Alternatively, they might reason that she will search in the left-hand box because she last saw it there. The exact same prediction would be made irrespective of whether or not infants encode her representation as false or not.

Is it plausible to think that infants might represent the agent's representation, without considering how that representation compares with reality? One argument might be that, due to their limited ability to act on the world themselves, the most important events to encode and retain are those that are relevant for others' actions. The location of the ball as represented by the agent

will be predictive of her actions, but the location of the ball as represented by the infant will not be. One possibility then is that infants maintain in working memory only the most relevant representation for their purposes. In non-verbal false-belief tasks, during habituation or familiarization trials, infants are essentially told that, what is relevant is what the agent will do. For example, in Southgate et al. (2007), infants observe two familiarization trials during which they are trained that when the windows light up and they hear this simultaneous chime sound, the agent will reach through one of the windows. Eliciting anticipatory saccades to the windows works only because the infant is motivated to care about this aspect of the event (either because they care anyway, or because we train them that this is what is relevant). Plenty of evidence shows that we do not always encode or retain information that we do not deem to be relevant, and that we pay attention to some things at the expense of others (e.g. Duncan, 2006). Although phenomena such as change or inattention blindness have been most extensively studied in adults, there is also evidence showing that infants encode or retain different aspects of events depending on what they perceive as being most relevant (e.g. Mareshal & Johnson, 2000; Yoon et al., 2008). Thus, it is not implausible to take seriously the possibility that infants might not encode, or might not retain, the true location of an object that is the target of someone else's actions.

Moreover, recent research suggests that sometimes, we may not store our own and others' perspectives as distinct versions of reality that we can keep separate and compare, and that our encoding of others' experiences can even interfere with our encoding of our own experience of an event. For example, in a task in which adults had to make rapid judgments about how many objects were present in a scene, Samson and colleagues found that reaction times were slower when the number of objects that were visible to the subject and an avatar were different, suggesting that the agent was also computing the scene from the others' perspective and had to overcome this representation in order to make a correct judgment (Samson, Apperly, Braithwaite & Andrews, 2010). An interesting aspect of these results was that adults were actually slightly faster to judge someone else's perspective than they were to judge their own perspective, raising the possibility that sometimes, someone else's perspective might be given greater importance than one's own. In a similar study, Kovacs and colleagues gave subjects an object-detection task (Kovacs, Teglas & Endress, 2010). Participants watched as a Smurf character rolled a ball behind an occluder. Sometimes the participant's and the agent's beliefs differed (i.e. the participant had seen the ball roll away while the agent was away, so the participant knew the object was absent, but the agent should think it is present) and sometimes they were the same (i.e. both saw the ball leave, or both saw the ball roll back behind the occluder). Kovacs et al. (2010) measured participant's reaction times to detecting the presence of the ball once the occluder was lowered and found that even when participants had seen the ball roll away, if the

agent should think the ball is present (because he was not there when the ball rolled away), participants were just as fast at detecting the object as when they themselves should think it is behind the occluder (because they had not seen it roll away). Both these results suggest that sometimes we may encode the others' perspective even at the expense of our own. Kovacs and colleagues (2010) (p. 14) also extended their paradigm to 7-month-old infants and found a similar phenomenon: infants who had observed a ball roll away (so should represent no ball behind the occluder), but who knew that the Smurf was absent when the ball rolled away (so the Smurf should represent the ball as still being behind the occluder), looked longer when the ball was revealed to be absent than when the infant and the Smurf had both seen the ball roll away. These results raise the possibility that infants may not be representing the other's perspective of an event as a separate version of reality that they could compare with their own, and that their own representation of the event does not take precedence. However, although these results tell us that infants encode the others' perspective, and that it interferes with their own representation of an event, they do not tell us whether sometimes infants might *preferentially* encode or retain the perspective of the other.

Nevertheless, this hypothesis would provide an alternative explanation for why infant's performance on false-belief tasks is not subject to an egocentric bias. If infants encode the other's perspective, but either fail to encode, or fail to maintain in memory, events that occurred during her absence (which would not alter, or be relevant to, her subsequent actions), then there would be no alternative reality to inhibit. Infants would preferentially retain the aspects of the event that are directly relevant for the other's expected action, and while this would constitute a metarepresentation, it would not amount to an understanding of belief as a concept which can be true or false, and the truth value of the others representation may not be considered.⁷ This kind of representation would suffice for infants to generate accurate predictions about what others will do, but it would also have limitations. Returning to Dennett's Punch and Judy example, infants might be able to understand that Punch wants to get rid of Judy and that Punch represents Judy as being in the box and so will push the box over the cliff. However, they would not find it funny in the way that 4-year-olds do because the humor of this event comes from the fact that children know that she's not really in the box, and they can compare their knowledge with Punch's belief, and it is this difference in representations that makes the event funny. Furthermore, while representing the others' perspective without considering its truth value would be sufficient for action predictions, it is likely to be insufficient for explaining why others are acting as they are. This kind of perspective-taking might enable you to predict where Sally will search, but you will not be able to explain why Sally comes up empty-handed (see Andrews, 2003 for a similar distinction).

Finally, if infants could pass false-belief tasks without understanding that the belief is false, it may provide an explanation for why children do not pass the standard false-belief task until around 4 years of age. In most non-verbal tasks, there is nothing that forces infants to consider the real state of affairs, or to consider the others' perspective as distinct from their own. Infants are free to attend to whatever they deem most relevant. On the other hand, in verbal false-belief tasks, instructions often make it explicit that the other person has a perspective that differs from the child's own perspective and experimenters typically highlight the reality by asking children memory questions like "where is the ball really?" or "what is in the box?". This kind of questioning may lead children to consider the real location of the object and, until they have the ability to deal appropriately with two conflicting representations, they may default to answering in terms of their own perspective.

Conclusions

Our view of infant's social cognitive abilities has undergone a radical transformation in the last decade with the publication of results suggesting that even very young infants are considering others' perspectives on the world, and are not, as was once thought, trapped in an egocentric viewpoint. The results of the non-verbal, and indirect elicited response, false-belief tasks suggest **(p.15)** that, long before children turn four, they understand that others' own experiences modulate their actions. The discovery that this understanding exists early in childhood not only gives support to previous interpretations of children's behavior in naturalistic settings (Dunn, 1991), but also provides a solution to the tension that has existed between the views that young children are competent communicators but incompetent mindreaders (Breheny, 2006).

However, while infants can track events from others' perspectives, this need not imply that the representations that underlie their mindreading abilities are the same as the representations that support belief reasoning in older children and adults. For example, theory of mind comprises not only the formation of a representation of someone's thought or perspective, but the process of using that representation to generate predictions about how those thoughts will influence behavior (Dennett, 1978). However, while looking-time data tell us that 7- and 10-month-olds are representing events from others' perspectives (Kovacs et al., 2010; Luo, 2011), they do not tell us whether these younger infants can use these inferences to generate predictions about how others will behave. Furthermore, we do not know whether infants, like adults, understand that a belief can be false, or when attributing a belief to someone, whether young infants are considering the truth value of that belief. Finally, while we readily talk about infants' understanding of mental state concepts like "knowledge" and "beliefs," we do not really know what infants mental state concepts might look like. Do infants possess distinct mental state concepts like knowledge and belief? It is possible that initially, infants may have available only a core epistemic-state concept (P. Carruthers, personal communication), which only gradually becomes

differentiated with development. The kind of nuanced mental state concepts that adults possess are likely to be culture- and language-dependent and, while there may exist core psychological concepts that are universal and available pre-linguistically, further mental state concepts may be dependent on linguistic input (Scholl & Leslie, 1999; Wellman, 1998). Thus, while the evidence suggests that infants are tracking others' epistemic states from an early age, it is important to recognize that there are various ways in which this understanding might differ from an adult-like understanding, while ensuring that infants are very good at figuring out what others will do.

Acknowledgements

I would like to thank Teodora Gliga for valuable discussion and comments on an earlier version of this chapter. VS is supported by a Wellcome Trust Research Career Development Fellowship (088427/Z/09/Z).

References

Bibliography references:

Andrews, K. (2003). Knowing mental states: The asymmetry of psychological prediction and explanation. In Q. Smith and A. Jokic (Eds), *Consciousness: New Philosophical Perspectives* (pp. 201–219). Oxford: Oxford University Press.

Baillargeon, R., Scott, R., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14, 110–18.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, 20, 191–208.

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

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

Breheny, R. (2006). Communication and folk psychology. *Mind and Language*, 21(1), 74–107. **(p.16)**

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-olds show false belief understanding in an active helping paradigm. *Cognition*, 112, 337–42.

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory control in young children's difficulties with deception and false belief. *Child Development*, 69, 672–91.

Carpendale, J., & Lewis, C. (2006). *How Children Develop Social Understanding*. Oxford: Blackwell.

Dennett, D. C. (1978). Beliefs about beliefs. *Brain and Behavioral Sciences*, 1, 568–70.

Duncan, J. (2006). Brain mechanisms of attention. *Quarterly Journal of Experimental Psychology*, 59, 2–27.

Dunn, J. (1991). Understanding others: Evidence from naturalistic studies of children. In A. Whiten (Ed.), *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading* (pp. 51–61). Oxford: Blackwell.

Friedman, O., & Petrashek, A. R. (2009). Children do not follow the rule “ignorance means getting it wrong”. *Journal of Experimental Child Psychology*, 102, 114–21.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences*, 7(7), 287–92.

Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56, 165–93.

German, T. P., & Leslie, A. M. (2000). Attending to and learning about mental states. In P. Mitchell & K. Riggs (Eds), *Children’s Reasoning and the Mind* (pp. 229–252). Hove: Psychology Press.

Gomez, J. C. (2008). The evolution of pretence: From intentional availability to intentional non-existence. *Mind & Language*, 23(5), 586–606.

Gopnik, A., & Astington, J. W. (1988). Children’s understanding of representational change and its relation to understanding of false belief and the appearance-reality distinction. *Child Development*, 59(1), 26–37.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Hedger, J. A., & Fabricius, W. V. (2011). True belief belies false belief: Recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development. *Review of Philosophy and Psychology*, 2(3), 429–447.

Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, 21, 101–14.

Knudsen, B., & Liszkowski, U. (2012a). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science*, 15(1), 113–122.

Knudsen, B., & Liszkowski, U. (2012b). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy*, 17(6), 672–691.

Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330(6012), 1830–4.

Leslie, A. M. (1994). ToMM, ToBy and Agency: Core architecture and domain specificity. In L. Hirschfeld and S. Gelman (Eds), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 119–148). New York: Cambridge University Press.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science*, 1, 247–53.

Low, J., & Wang, B. (2011). On the long road to mentalism in children's spontaneous false-belief understanding: are we there yet? *Review of Philosophy and Psychology*, 2(3), 411–28.

Luo, Y. (2011). Do 10-month-old infants understand others' false beliefs? *Cognition*, 121(3), 289–98.

Mareschal, D. & Johnson, M. H. (2003). The “what” and “where” of object representations in infancy. *Cognition*, 88, 259–76.

Meltzoff, A. M., & Brookes, R. (2008). Self-experience as a mechanism for learning about others: A training study in social cognition. *Developmental Psychology*, 44(5), 1257–65.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 255–8. **(p.17)**

Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a “theory of mind”. *Philosophical Transactions of the Royal Society, B*, 362, 731–44.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5(2), 125–37.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, 308, 214–16.

Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Basic.

Povinelli, D. J., & Vonk, J. (2003). Chimpanzee minds: suspiciously human? *Trends in Cognitive Sciences*, 7(4), 157–60.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4, 515–26.

Ruffman, T. (1996). Do children understand the mind by means of simulation or a theory? Evidence from their understanding of inference. *Mind & Language*, 11, 388–414.

Ruffman, T., & Perner, J. (2005). Do infants really understand false belief? Response to Leslie. *Trends in Cognitive Sciences*, 9, 462–3.

Ruffman, T., Slade, L., Redman, J. (2005). Young infants' expectations about hidden objects. *Cognition*, 97, B35–43.

Ruffman, T., Taumoepeau, M., & Perkins, C. (2011). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, 30(1), 87–104.

Samson, D., & Apperly, I. A. (2010). There is more to mind reading than having theory of mind concepts: New directions in theory of mind research. *Infant and Child Development*, 19, 443–54.

Samson, D., Apperly, I. A., Braithwaite, J., & Andrews, B. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1255–66.

Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and “theory of mind”. *Mind & Language*, 14, 131–53.

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, 80, 1172–96.

Scott, R. M., Baillargeon, R., Song, H., & Leslie, A. M. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, 61(4), 366–95.

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science*, 22, 878–80.

Song, H., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology*, 44(6), 1789–95.

Song, H., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition*, 109, 295–315.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, 16, 907–912.

Southgate, V., & Csibra, G. (2009). Inferring the outcome of an ongoing novel action at 13 months. *Developmental Psychology*, 45, 1794–8.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18, 587–92.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18(7), 580–6.

Surian, L., & Leslie, A. M. (1999). Competence and performance in false belief understanding: A comparison of autistic and normal 3-year-old children. *British Journal of Developmental Psychology*, 17, 141–55.

Trauble, B., Marinovic, V., & Pauen, S. (2010). Early theory of mind competencies: Do infants understand others' belief? *Infancy*, 15, 434–44. **(p.18)**

Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, 9, 217–32.

Wellman, H. M. (1998). Culture, variation and levels of analysis in folk psychologies. *Psychological Bulletin*, 124, 33–6.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development*, 72, 655–84.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in children's understanding of deception. *Cognition*, 13, 103–28.

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

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development*, 22(2), 145–60.

Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 13690–5.

Yott, J., & Poulin-Dubois, D. (2012). Breaking the rules: Do infants have a true understanding of false belief? *British Journal of Developmental Psychology*, 30(1), 156–71.

Notes:

- (1) See Perner (2010) for arguments against this position.
- (2) Both the transparent and opaque blindfold looked identical when worn by someone else, but regardless of condition, all infants saw a movie in which the agent wore the opaque blindfold.
- (3) See Penn & Povinelli (2007) for further arguments as to why expectations modulated by self-experience necessarily entails mental state attributions.
- (4) It is possible that, if understood, the initial phrase “do you remember what I put in here?” could have cued infants to think about which object the experimenter put in there, rather than about what the experimenter thinks is in there. Thus, in an additional experiment, we used a different phrase (“do you know what is in here”) with identical results.
- (5) As opposed to a gaze-based (looking-time or anticipatory-looking) response, which many have termed an “implicit” response. However, it should be recognized that this is simply a description of the two kinds of responses and does not tell us anything about the computations involved.
- (6) However, as we will discuss in the following section, inhibition may be required to a greater extent on tasks in which the infants own perspective is explicitly highlighted.
- (7) Samson & Apperly advocate a similar position.