Developmental Science

Developmental Science 17:5 (2014), pp 647–659 DOI: 10.1111/desc.12148

TARGET ARTICLE WITH COMMENTARY AND RESPONSE

False belief in infancy: a fresh look

Cecilia Heyes

All Souls College & Department of Experimental Psychology, University of Oxford, UK

Abstract

Can infants appreciate that others have false beliefs? Do they have a theory of mind? In this article I provide a detailed review of more than 20 experiments that have addressed these questions, and offered an affirmative answer, using nonverbal 'violation of expectation' and 'anticipatory looking' procedures. Although many of these experiments are both elegant and ingenious, I argue that their results can be explained by the operation of domain-general processes and in terms of 'low-level novelty'. This hypothesis suggests that the infants' looking behaviour is a function of the degree to which the observed (perceptual novelty) and remembered or expected (imaginal novelty) low-level properties of the test stimuli – their colours, shapes and movements – are novel with respect to events encoded by the infants earlier in the experiment. If the low-level novelty hypothesis is correct, research on false belief in infancy currently falls short of demonstrating that infants have even an implicit theory of mind. However, I suggest that the use of two experimental strategies – inanimate control procedures, and self-informed belief induction – could be used in combination with existing methods to bring us much closer to understanding the evolutionary and developmental origins of theory of mind.

Introduction

Do infants understand that other agents can have false beliefs? Following the ground-breaking work of Onishi and Bailargeon in 2005, this important question has been given a positive answer by more than 20 experiments examining infants' looking behaviour (Onishi & Baillargeon, 2005). The standard interpretation of their results has been questioned incisively (Apperly & Butterfill, 2009; De Bruin & Newen, 2012; Perner, 2010; Perner & Ruffman, 2005; Ruffman, Taumoepeau & Perkins, 2012), but the data from these infant false belief (FB) studies are still well on their way to establishing a new consensus in developmental science; to persuading us that the human capacity to attribute mental states – 'theory of mind', 'mentalizing' or 'mind reading' – is largely inborn (Carruthers, 2013). I argue that this conclusion would be premature.

This article is unlike previous critical analyses in three respects. First, it is comprehensive, offering a detailed and unified account of nearly all experiments published to date reporting positive evidence of false belief attribution in children below 2 years of age. Second, it offers a novel perspective rooted in cognitive science. Previous critiques have proposed that domain-specific 'behaviour rules' are responsible for infants' behaviour in FB tests, but I suggest that domain-general processes could be responsible for the infants' behaviour. Finally, I am optimistic. Some researchers state or imply that there could never be adequate empirical evidence for the hypothesis that non-linguistic creatures – preverbal infants and animals – can attribute FBs, and suggest that we must rely on appeals to parsimony, rather than experiments, to resolve the issue. In contrast, I argue that the problem of distinguishing 'mind reading' from 'behaviour reading' in agents without language is a difficult but ultimately tractable scientific problem, and suggest some experimental strategies that could be helpful.

The first section of the article is devoted to a detailed review of experiments testing for FB attribution in infants. The second discusses new methods for testing the hypothesis that infants can attribute FBs.

Empirical review: the devil in the details

Let's begin with a broad summary of what happens in most infant FB experiments: Infants are presented with a sequence of events – occurring live in a puppet theatre or on video – involving what an adult would understand to be an agent (human or nonhuman), one or more focal

Address for correspondence: Cecilia Heyes, All Souls College, Oxford OX1 4AL, UK; e-mail: Cecilia.Heyes@all-souls.ox.ac.uk
For a commentary on this article see Scott and Baillargeon (2014).

objects (e.g. toys), and ancillary objects (e.g. containers, occluders). Part of the sequence is intended by the researcher(s) to make the infants attribute to the agent a true or a FB about the location of a focal object in relation to ancillary objects (e.g. that a particular toy is in the green box rather than the yellow box). The stimulus sequence terminates with an act performed by the agent that the researchers expect to surprise, or not to surprise, the infants, depending on whether the researchers think it likely that the infants have attributed a true or a FB to the agent. If infants spend longer looking at the events that the researchers expect to be surprising than at the events the researchers do not expect to be surprising, then the experiment is interpreted as providing evidence of FB attribution in infants.

This summary is unusual because it acknowledges that infant FB experiments involve three, rather than two, levels of intentionality. As a reader who wants to understand the experimental logic, it is not enough for me to keep track of what the infants may or may not believe about the agent's beliefs. I must also take into account what the researchers may or may not believe about what the infants may or may not believe about the agent's beliefs. This is demanding inferential work. Therefore, like the authors of infant FB studies, I'll usually describe the experimental procedures from the perspective of an adult with mature, explicit theory of mind; for example, saying that 'the agent reached towards the green box', rather than 'the arm-shape moved in an oblique, minimum jerk trajectory towards the green cube-shape'. It would be hopelessly obscure to use more neutral language all the time. But we should not forget that it is in an important sense the purpose of these experiments to find out whether infants appreciate that the 'arm-shape' is attached, not merely to a human body (Slaughter, Heron-Delaney, Christie, Slaughter & Brownell, 2011), but to a fully-fledged agent.

In the sections that follow I argue that the results of infant FB studies are due to *low-level novelty.* In many cases, the novelty depends only on the observable properties of the test stimuli; infants look longer at test events that, when compared with events encoded earlier in the experiment, display new spatiotemporal relations among colours, shapes and movements (*perceptual novelty*). In other cases, the novelty depends on imagination as well as the manifest characteristics of the test stimuli; infants look longer at test events that, when compared with events encoded earlier in the experiment, are remembered or imagined to have new spatiotemporal relations among colours, shapes and movements (*imaginal novelty*). The low-level novelty account assumes that domain-general processes of perception, attention, motivation, learning and memory are solely responsible for encoding the events that infants witness in FB experiments and for modulating the infant's looking behaviour. It is not rooted in particular theories of how these domain-general processes operate. For example, the learning component could be 'statistical', 'associative', 'Bayesian' and/or 'inferential'. However, the low-level novelty account does draw on what is known about the way in which these domain-general processes operate; it draws on information, derived from experiments using predominantly asocial stimuli and a broad range of species, about the variables affecting phenomena such as attentional capture (Ruz & Lupiáñez Castillo, 2002), habituation (Rankin, Abrams, Barry, Bhatnagar, Clayton, Colombo, Coppola, Geyer, Glanzman, Marsland, McSweeney, Wilson, Wu & Thompson, 2009), stimulus generalization (Pearce, 1987), and retroactive interference.

For example, retroactive interference – disruption of memory for event X because it is followed by event Y (Pearce, 2008) – is a robust phenomenon found in a broad range of nonhuman species (including monkeys (Fuster & Bauer, 1974) and pigeons (Grant, 1988)), and in human adults (Yoon, Curtis & D'Esposito, 2006) and infants (Diamond, 1985; Marcovitch & Zelazo, 2003). In infants, retroactive interference contributes to the A-not-B error; in experimental paradigms similar to those used to test for FB attribution in infants, a distracting event, occurring after the infant has seen an object hidden at location B, disrupts memory for this event and thereby encourages her to search at location A (Diamond, 1990; Longo & Bertenthal, 2006). Resistance to retroactive interference – maintenance of an active representation of X in spite of distraction by Y – depends on prefrontal cortex, and is therefore likely to be particularly weak in infancy (Miller, Erickson & Desimone, 1996; Olesen, Macoveanu, Tegnér & Klingberg, 2007).

Like FB explanations, the low-level novelty account assumes that infants' looking behaviour in FB experiments reflects what can loosely be described as the extent to which the test event violated expectations or was surprising. However, the low-level account suggests that the novelty generating the surprise is present at a lower representational level; at a level where the events witnessed by the infants are represented as colours, shapes and movements, rather than as actions on objects by agents.

'Belief induction' variables play a crucial role in infant FB research. These are experimental manipulations designed to make infants attribute a true or FB to the agent about the properties of an object. I will argue that belief induction variables work – they have an impact on infants' looking behaviour – not by changing the beliefs that infants attribute to others, but by changing what the infants themselves believe about the properties of experimental objects; they influence first-order, rather than second-order, beliefs.

Three belief induction variables have been used in experiments to date, each relating to the disposition of the agent during key events, i.e. movements of the focal and/or ancillary objects that are likely to influence the content of the agent's beliefs: (1) Most commonly the agent is visible to the infant (present) or not visible to the infant (absent) during key events. Researchers assume that infants are likely to ascribe a true belief to the agent in the present condition and a FB in the absent condition. In the review that follows I'll point out that the present/absent variable is typically implemented in a way that is likely to cause retroactive interference. Consequently, when the agent is not visible during key events, the infant herself is less likely to remember those events. (2) In some studies the infant can see that there is (obstacle) or is not (no obstacle) an opaque object located between the agent and key events. In this case, researchers assume that infants are likely to attribute a true belief in the no obstacle condition and a FB in the obstacle condition. However, these obstacle studies have confounded the agent's view and the infant's view of key events. (3) In several studies, the agent's front (facing) or back (back turned) was visible to the infant during key events, and it was assumed that the infant was likely to attribute a true belief in the facing condition and a FB in the back turned condition. However, in these studies back turning involved or was accompanied by a stimulus that is likely to have distracted the infant's own attention from key events.

The empirical review is divided into two parts. The first discusses studies in which perceptual novelty is sufficient to explain the experimental results, and the second examines studies showing that imaginal novelty also plays a role.

Do we really need such a detailed review of infant FB experiments? I've tried hard to convince myself that we do not. As I grappled with each experiment – trying to assimilate the methodological details, to imagine the infant's experience, and to resist the intuitive pull of FB interpretations – I've tried to persuade myself it would be enough to say that the results of infant FB experiments can be explained by low-level novelty, and to illustrate the point with a few carefully chosen examples. But that kind of 'cherry picking' really wouldn't do (Baillargeon, 1999). To be useful and interesting, an alternative account must explain the full range of phenomena embraced by the original account, and do it in an equally unified way. It's easy but unhelpful to offer a different alternative explanation for each experimental result. So, I needed to check that none of the experiments resist a low-level novelty interpretation, and I assume many readers will want to do the same. (In that case, the reader will need to refer to the extensive notes in the Supplementary Information, each indicated by a superscript number in the text that follows.) In the second part of the article, the focus will broaden again as we consider new experimental strategies.

Perceptual novelty

Green and yellow boxes

Onishi and Baillargeon's (2005) original, nonverbal test for FB attribution in 15-month-old infants established a model that has been adapted in various ways by Baillargeon's group and others. Their procedure had three phases: familiarization, belief induction and test (see Figure 1). In the familiarization phase, the infant saw a similar sequence of stimulus configurations on three successive occasions. On the first occasion, a human agent manipulated a toy before reaching towards a green box on the infant's right, and away from a yellow box on the infant's left, and placing the toy in the green box. In the second and third trials, the agent simply reached towards the green box. In the final test phase, the infant saw the same sequence that had been presented during the familiarization trials (the agent reaching towards the green box; Group Green), or an alternative sequence (the agent reaching towards the yellow box; Group Yellow). Whether Group Green or Group Yellow looked more at their test sequence varied according to the events presented in the second, belief induction phase. Group Yellow looked longer than Group Green – implying that they were more surprised than Group Green – when the belief induction sequence was a shuffling movement of the yellow box with the agent present (true belief-green or 'TB-green' condition), and when it was ghostly (apparently self-propelled) movement of the toy from the green to the yellow box with the agent absent (FB-green condition). In contrast, Group Green looked longer than Group Yellow when the belief induction sequence was ghostly movement of the toy from green to yellow with the agent present (TB-yellow condition), and when it had two components – ghostly movement of the toy from green to yellow with the agent present, followed by ghostly movement back to green with the agent absent (FB-yellow condition).

The names of the conditions (TB-yellow, FB-green, etc.), which I've transcribed from Onishi and Baillargeon's (2005) article, give a good guide to the standard interpretation of these results. It assumes that the infants expected, on the basis of the familiarization trials, that the agent would reach towards the box where the agent

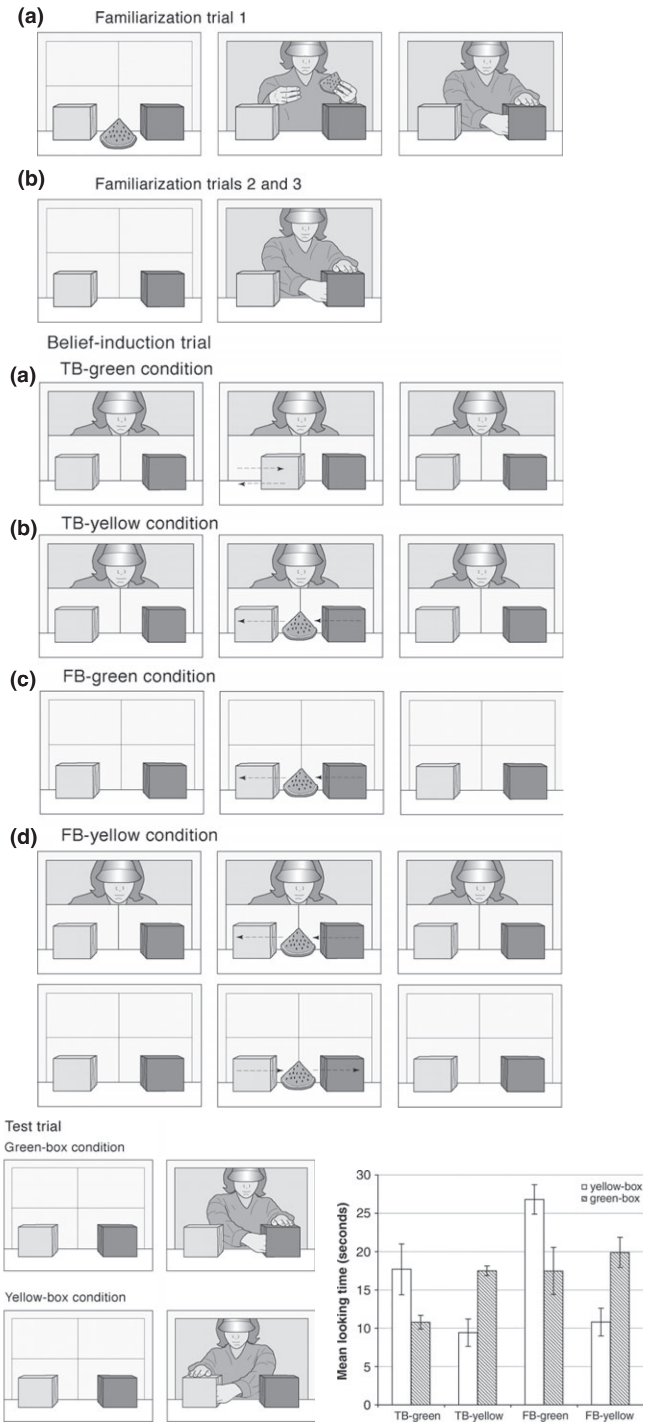

Figure 1 *The procedure and results of Onishi and Baillargeon's (2005) ground-breaking study of false belief in infancy. Reprinted with permission.*

believed the toy was located, and, on the basis of their experience in the belief induction trial, that the infants attributed to the agent a true or a FB that the toy was in the yellow box or in the green box. Thus, for example, the standard interpretation says that Group Yellow looked longer than Group Green in the FB-green condition because the infants in this condition attributed to the agent a FB that the toy was in the green box, and were therefore more surprised when she reached for the yellow than when she reached for the green box. Why would the infants in the FB-green condition attribute to the agent a FB that the toy was in the green box? Because the infants believed, among other things, (1) that the agent was not merely invisible but absent when the toy moved from the green to the yellow box; (2) that an agent cannot see an event that occurs in her absence; (3) that an agent who has not seen an object relocated is unlikely to have a true belief about the object's new location; and (4) that an agent who was absent when an object was moved is likely to believe that the object is still at the last location where it was seen by the agent. A great deal more would need to be said to specify all that the standard interpretation assumes about infants' capacity to form first-order and second-order beliefs on the basis of visual experience. However, this list should be sufficient to remind us that, to attribute a FB, an infant would need to know – consciously or unconsciously, implicitly or explicitly – a great many things about objects, minds and how they interact.

Here's the alternative, low-level novelty interpretation of the results from Onishi and Baillargeon's experiment. In the TB-green and FB-green conditions the infants in Group Yellow looked for longer than those in Group Green because the event presented to Group Yellow was more perceptually novel, relative to the infants' familiarization experience, than the event presented to Group Green. Recall that all of the infants had seen the agent-shape move towards green three times during familiarization, but none of them had previously seen it move towards yellow. The infants in the TB-yellow and FB-yellow conditions showed the opposite effect – Group Green looked for longer than Group Yellow – because, after familiarization and before the test, these infants saw an event (movement of the toy-shape towards yellow), that was visually similar to the yellow test event (movement of the agent-shape towards yellow), and therefore reduced the novelty of the yellow test event. The FB-green infants also saw movement of the toy towards yellow during belief induction but their memory for this event was impaired because it was immediately followed by a salient distractor event – the unexpected reappearance of the agent at the beginning of the test phase. Therefore, due to retroactive interference, it did not reduce the perceptual novelty of the yellow event.[1]

This low-level novelty interpretation of Onishi and Baillargeon's results is consistent with the findings of subsequent studies in which infants were tested in one of their TB (Poulin-Dubois & Chow, 2009)[2] or FB conditions (Yott & Poulin-Dubois, 2012)[3] after a period of pretraining.

Agent Smurf

In a more recent study using the presence or absence of an agent as the belief induction variable, the agent was a Smurf and the stimuli were presented to 7-month-old infants on video (Kovács, Téglás & Endress, 2010). In two familiarization trials, the Smurf was visible as a ball rolled behind an occluder, emerged on the right of the computer screen (right move), and rolled back behind the occluder (left move). At this point the occluder fell forwards, revealing the ball. On test, again with the agent visible, the screen fell forwards but there was nothing behind it. Immediately before the test, the infants saw a belief induction sequence that varied by condition. In the Familiar-Present condition (labelled P+A+ by Kovács *et al.*, 2010) the infants saw a sequence of ball movements similar to those presented in familiarization trials – a right move followed by a left move – with the agent present throughout. In the Familiar-Absent condition (P+A−), the agent was present during the right move, but absent during the left move. In the other two conditions the infants saw a right move followed by a left move followed by another right move, with the agent present (Novel-Present, P−A−) or absent (Novel-Absent, P−A+).

These conditions were compared within subjects in a carefully controlled pairwise fashion. The first experiment (Experiment 4 in Kovács *et al.*, 2010) showed that infants looked for longer at the test event (occluder falls, no ball) in the Familiar-Present than in the Novel-Present condition. The low-level novelty account suggests that this was because the Familiar-Present belief induction sequence (right–left) was more like the familiarization sequence (right–left) than was the Novel-Present belief induction sequence (right–left–right). Therefore, the Familiar-Present treatment set up a stronger expectation that the belief-induction trial would end in the same way as the familiarization trials, i.e. with the appearance of a ball behind the occluder. In the second experiment (Experiment 5) infants looked longer at the test event in the Novel-Absent condition than in the Novel-Present condition. In both of these conditions, the novel right–left–right sequence occurred during belief induction. However, due to retroactive interference, memory for the latter part of this sequence – the part that made it different from the familiarization sequence – is likely to have been impaired in the Novel-Absent condition by the reappearance of the agent at the end of the sequence, i.e. in the test trial. Therefore, the belief induction events encoded in the Novel-Absent condition were more like those of the familiarization trials than the belief induction events encoded in the Novel-Present conditions, leading to a stronger expectation that the belief induction sequence would end in the same way as the familiarization sequence, with the sight of a ball.

This low-level interpretation is also consistent with the results of two control experiments reported by Kovács *et al.* (2010).[4]

The caterpillar's eye view

Many FB studies make some use of occluders, but in certain studies they play a crucial role; they replace presence or absence of the agent as the primary means by which the researcher seeks to manipulate whether infants attribute a true or FB to the agent. Occluders played this crucial role in a study where 13-month-old infants watched videos involving a caterpillar agent (Surian, Caldi & Sperber, 2007) (Experiment 1). During the familiarization phase, the infants saw a stage with two occluders. The one on the left partially obscured the infant's view of a piece of cheese, and the one on the right partially obscured the infant's view of an apple.[5] In each of four familiarization trials, the caterpillar moved up the centre of the computer screen and then turned left towards the cheese. In the fifth familiarization trial, the infant saw a human hand place the cheese behind the right occluder and the apple behind the left occluder; that is, in the reverse positions relative to their locations in previous trials. In a single test trial, the infants saw the caterpillar move up the centre of the screen and turn right towards the cheese (New Path – Old Goal) or turn left towards the apple (Old Path – New Goal), with short occluders, affording the infant a full view of the caterpillar and objects (TB condition) or tall occluders, affording the infant a partial view of the agent and objects (FB condition).[6] The results indicated that the infants looked longer at the Old Path – New Goal than at the New Path – Old Goal in the TB condition, and looked for roughly equal time at these two events in the in the FB condition.

These results can be explained in a straightforward way by low-level perceptual novelty. In the TB test, the caterpillar's path (a briefly presented left or right turn relative to the occluders) was less salient than his goal (cheese or apple) because the occluders were small and the objects were fully visible to the infants. Therefore, they registered more surprise when the caterpillar moved towards a novel object relative to familiarization, than when he took a novel path relative to familiarization.

However, in the FB condition, when the occluders were high and the objects were only partially visible to the infants, the caterpillar's path and destination object were roughly equally salient to the infants, and therefore the two test events were roughly equally (un)surprising. Thus, the occluders in this experiment by Surian *et al.* (2007) were intended to manipulate the beliefs attributed by the infants to the agent. However, because the agent's view was confounded with the infant's view, it is at least equally likely that the occluders influenced the infants' behaviour via their impact on the infants' beliefs about events, rather than the infants' beliefs about the agent's beliefs about events. The results of another study, which used occluders and manipulated an agent's path and goal, can also be explained in terms of low-level perceptual novelty (Luo, 2011a).[7]

In summary: Many of the looking time results that are understood to show that infants can attribute FBs – including those of the original study by Onishi and Baillargeon (2005) – can be explained in terms of low-level perceptual novelty; by assuming that looking time was a function of the degree to which the observable, low-level features of the test event differed from those encoded by the infants earlier in the experiment. These results are explicable by low-level perceptual novelty because the present/absent belief induction variable was implemented in a way that is likely to have caused retroactive interference, and the obstacle/no obstacle belief induction variable was confounded with factors affecting the infants' own perceptual access to key events.

Imaginal novelty

The studies reviewed in this section indicate that, although it is low-level novelty that generates looking time effects in infant FB experiments, this novelty does not necessarily depend only on the manifest or observable features of the test events; it can also depend on low-level features that the infants associate with the test objects on the basis of their experience earlier in the experiment. One could say that it depends on what the infants believe about the low-level features of the test objects. For example, we could say that, after seeing a toy enter an opaque box, the infant believes that the box contains the toy, and will be surprised if the agent reaches for it on test after reaching repeatedly for a different toy earlier in the experiment. However, it is no more necessary to talk about beliefs in relation to imaginal than perceptual novelty. All that is required to produce an imaginal novelty effect is that the sight of the box-shape or box-colour activates a representation of the low-level properties of another object, such as the shape or colour of a particular toy. To put it more vividly, the infant needs only to have a mental picture – as if the box had transparent walls – of the toy with the box. The important thing to note about imaginal novelty is that, in contrast with FB explanations, it involves the infant ascribing low-level features (colours, shapes, sounds) to objects, rather than high-level features (true and FBs) to agents.

I have categorized infant FB studies in terms of perceptual and imaginal novelty to highlight that some studies do, and others do not, require us to consider how infants represent the test objects on the basis of their past experience. Typically, studies demonstrating imaginal novelty trigger this requirement by increasing the perceptual distinctiveness of the alternative objects in the belief inductive phase relative to the test phase. My categorization of studies as demonstrating perceptual or imaginal novelty does not imply that fundamentally different processes are responsible for infant behaviour in the two types of studies, and certainly not that different processes are strategically deployed by the infants. The low-level novelty account assumes that infants' reactions to test events always depend on how they represent those events, and that this, in turn, depends on a combination of incoming stimulation and past experience. The effects of past experience are more likely to become evident in behaviour when the events are less distinctive in the former respect than in the latter.

The caterpillar comes and goes

In the previous section I discussed the first of two experiments reported by Surian *et al.* (2007), which can be explained in terms of perceptual novelty. In their second experiment, which provided evidence of imaginal novelty, they reverted to using the presence or absence of the agent as a belief induction variable. This experiment was the same as the previous one except that (1) all infants had test trials with tall occluders, and (2) for half of the infants the agent was visible when the locations of the objects were reversed in the fifth familiarization trial (TB condition), while for the other half the agent was invisible during this reversal (FB condition). The results were similar to those of Experiment 1: infants looked longer at the Old Path – New Goal than at the New Path – Old Goal in the TB condition, and looked for roughly equal time at these two events in the FB condition. In this experiment, as in previous experiments we have examined, presence versus absence of the agent in the belief induction phase (i.e. the fifth familiarization trial) was confounded with continuity versus abrupt change between belief induction and the test trial; in the FB condition but not in the TB condition, the agent reappeared just after belief induction. If this significant, unexpected event caused retroactive interference in the FB condition, the infants in the TB condition will have had a stronger representation of the locations of the objects during the test trial (e.g. a clearer mental picture of the cheese behind the right occluder), leading them to be more surprised when the agent took an old path to a new goal, rather than a new path to an old goal.

Blue hair

A role for imaginal novelty is also indicated by an experiment that used the present versus absent belief induction variable to investigate 14.5-month-old infants' reasoning about false perceptions (Song & Baillargeon, 2008). During the familiarization phase, infants saw an agent reach for a doll with blue hair and away from a toy skunk on four occasions. In the test trial, they saw the same agent reach towards a 'plain box' or a box of the same colour and shape with a tuft of the doll's blue hair attached ('hair box'). In the intervening belief induction sequence, the infants saw the doll placed in the plain box and the skunk in the hair box, when the agent was present/visible (TB condition) or when the agent was absent/invisible (FB condition). Infants in the TB condition looked for longer when the agent reached for the hair box, where the skunk was located, than when the agent reached for the plain box, where the doll was located. The infants in the FB condition showed the reverse effect.

To explain these results we need to take into account what the infants imagined about the test objects, and how those images made the test events like or unlike events encoded earlier in the experiment. Thus, as a result of encoding the events presented in the belief-induction trial, the TB infants are likely to have imagined, as if the box walls were transparent, the plain test box as the doll-in-plain box, and the hair test box as the skunk-in-hair box. In that case, the test event involving movement towards the skunk-in-hair box would be more novel relative to the familiarization trials (movement towards the doll and away from the skunk), and therefore attract more looking, than the test event involving movement towards the doll-in-plain box. The skunk-in-hair box had blue hair in common with the familiarization target, but otherwise it was completely different (skunk colours and shapes rather than doll colours and shapes). What about the infants in the FB condition? If we note, as we did with the previous experiments, that reappearance of the agent in the test trial is likely to have disrupted memory for events immediately preceding the reappearance, then the FB infants will not have imagined, or imagined vividly, the doll in the plain box and the skunk in the hair box. On test they simply saw movement towards a box with blue hair or an identical box without blue hair. The latter was more perceptually novel because movement towards blue hair was observed in the familiarization trials.

Imaginal novelty can also explain the results of studies using the logical equivalent of Onishi and Baillargeon's (2005) FB-green condition (Song, Onishi, Baillargeon & Fisher, 2008),[8] examining beliefs about object identity rather than location (Scott & Baillargeon, 2009)[9] and involving auditory cues to object location (Scott, Baillargeon, Song & Leslie, 2010).[10]

Bell ringing

An innovative study of 25-month-old infants (Southgate, Senju & Csibra, 2007) differed from those reviewed above in two respects. First, infants were encouraged to attribute a FB to an agent not by making the agent invisible during critical events, but by showing the infants that the agent's back was turned during those events. Second, like an earlier study of older children (35–53 months) (Clements & Perner, 1994), Southgate and colleagues used an 'anticipatory looking' rather than a 'violation of expectation' procedure. Rather than presenting alternative actions on test, and measuring infants' looking time as an index of how surprised they were by each action, they measured the direction and duration of eye movements at a time when, on the basis of previous trials, infants should be expecting the agent to initiate one of two actions. This method measures what infants expect to happen rather than the extent to which they are surprised by events already in progress.

In each of the video stimulus sequences used by Southgate *et al.* a human agent was seated behind a panel containing two windows, on the infant's left and right, and below each window was an opaque box. The agent's head was visible at the top of the panel. At the beginning of each familiarization trial, a puppet appeared at the bottom of the screen and placed a ball in one of the two boxes; in the left box in the first familiarization trial, and in the right box in the second. When the puppet had disappeared, both windows were illuminated and a tone sounded. 1.75 seconds later, a hand came through the window above the box containing the ball, moved towards that box and retrieved the ball. The belief induction trial always began with the puppet placing the ball in the left box. In the FB1 condition, the puppet then moved the ball to the right box and disappeared. At this point, a bell began to ring, and the agent turned her head away (looking up, behind, and to the infant's left). While the agent's head was turned away, the puppet reappeared, took the ball out of the right box and removed it from the scene. Once the puppet had disappeared, the bell stopped ringing, the agent turned back so that she was looking down at the boxes, and the test trial started, i.e. the windows were illuminated and the tone sounded. In the FB2 condition, the puppet disappeared immediately after placing the ball in the left box, and then the agent turned away (looking up, behind and to the infant's right). While her head was turned, the puppet moved the ball from the left to the right box, retrieved it from the right box, and removed it from the scene. Then the test trial started. The results showed that the infants in the FB1 condition tended to make their first eye movement after window illumination to the right window, and looked longer at the right than the left window, whereas infants in the FB2 condition made their first eye movement towards, and looked longer at, the left than the right window.

These results may indicate that the FB1 infants thought the agent believed that the ball was in the right box, whereas the FB2 infants thought the agent believed the ball was in the left box. Alternatively, it is possible that the FB1 infants looked to the right and the FB2 infants looked to the left simply because they, the infants, imagined the ball was on the right and on the left, respectively. That is, the bell ringing and head turning, that was supposed to signal to the infants that the agent could not see movements of the ball, may instead have distracted the infants so that they didn't see, or didn't remember, those movements. In that case, the infants might assume the ball to be at the location where they last saw it and, on the basis of their familiarization experience, expect the hand to appear above that location.[11]

The results of two other anticipatory looking studies are subject to the same low-level interpretation. One of these used a procedure very similar to that of Southgate and colleagues (Träuble, Marinović & Pauen, 2010),[12] and the other used a triangle as the 'agent' (Meristo, Morgan, Geraci, Iozzi, Hjelmquist, Surian & Siegal, 2012; Surian & Geraci, 2012).[13]

In summary: A substantial set of violation of expectation and anticipatory looking studies indicate that looking behaviour in infant FB experiments does not depend exclusively on the manifest characteristics of the test stimuli. In some cases, the novelty or 'surprise value' of the test events depends on properties they are imagined (remembered or expected) to have on the basis of previous experience within the experiment. However, like the findings reviewed in the previous section on perceptual novelty, the results examined in this section on imaginal novelty can be explained in terms of domain-general mechanisms that track low-level novelty; the degree to which the manifest and imagined low-level features of the test events – their colours, shapes and movements – differed from those encoded earlier in the experiment.[14]

Methodological recommendations: where to next?

The preceding review suggests that the FB and low-level novelty hypotheses are tied; they have equal scores in the contest to explain infants' looking behaviour in tests for FB attribution. The results of each study can be plausibly explained with reference to domain-specific mechanisms mediating FB attribution, and with reference to domain-general mechanisms sensitive to low-level novelty. Furthermore, the two hypotheses offer structurally similar and equally unified explanations. Each has a single, central construct – FB attribution and low-level novelty, respectively – and each makes additional assumptions in order to explain the results of experiments using different belief induction variables. For example, to explain effects of the present versus absent variable, the FB account assumes that infants believe that absent agents cannot see and therefore do not form true beliefs about key events, whereas the low-level account assumes that the reappearance of the agent causes retroactive interference. Finally on this scoreboard, each account has the benefit of convergent evidence; of coherence with the way in which other, similar data have been interpreted. For example, supporters of the FB hypothesis could point out that studies of helping behaviour seem to converge on the same conclusion as studies of looking behaviour (Buttelmann, Carpenter & Tomasello, 2009; Southgate, Chevallier & Csibra, 2010),[14] and supporters of the low-level hypothesis can cite convergent evidence from studies of the A-not-B error (Diamond, 1990). Similarly, but looking further afield for convergence, supporters of the FB hypothesis can point out that their assumptions are consistent with rich interpretations of young infants' reactions to object occlusion, discrimination of intentional from non-intentional actions, and imitation (Csibra & Gergely, 2011; Gergely, Nádasdy, Csibra & Bíró, 1995; Luo, 2011b; Woodward, 2009), whereas advocates of the low-level account can cite lean interpretations of the same abilities (Daum, Attig, Gunawan, Prinz & Gredebäck, 2012; Haith, 1998; Paulus, Hunnius, van Wijngaarden, Vrins, van Rooij & Bekkering, 2011a; Paulus, Hunnius, Vissers & Bekkering, 2011b; Sirois & Jackson, 2007).

There are a number of ways in which researchers could respond to a tie of this kind: (1) Each of us could stand by the account that we find, as a matter of intellectual taste, to be more congenial. This option is tempting, but not defensible. (2) It could be argued that, in spite of the tie, one of the accounts can be rationally preferred over the other because it is 'simpler' or more 'parsimonious'. This route is fraught with difficulty because there are many different kinds of parsimony, it is likely that the FB hypothesis would 'win' for some kinds (e.g. ease of use) and 'lose' for others (e.g. uniformity), and it's not clear how, if it all, the different types of parsimony can be weighted against each another (Fitzpatrick, 2008; Heyes, 1998, 2012). (3) It could be objected that, rather than being competitors, the FB and low-level accounts are equivalent; the low-level account merely describes, in an alternative terminology, what it is for an infant to have implicit theory of mind. Finally, we could accept that the FB and low-level accounts are competitors, and (4) conclude (possibly in combination with 1 or 2) that the question whether infants can attribute FB cannot be resolved by empirical means, or (5) use new experimental strategies in an attempt to break the tie. This section argues that the last option is the most promising, and suggests some experimental strategies that could be useful.

The third option is available because, in principle, one could define 'implicit theory of mind' in an extremely liberal way, as *any* set of psychological processes that under *some* circumstances align an agent's behaviour with the mental states of social partners; that is, co-vary with the mental state ascriptions to those social partners that would be made by a typical adult user of explicit theory of mind. Under this extremely liberal definition, even the kind of domain-general mechanisms that make humans and other animals respond to low-level novelty would count as an implicit theory of mind. However, the extreme liberality of this definition would also entail that a very wide range of species – including all animals that show habituation and dishabituation – also have an implicit theory of mind. If we want to pursue the hunch that theory of mind, even the implicit variety, has something to do with the complexity of human social behaviour relative to that of all or most other animals, we need to put some restrictions on what would count as implicit theory of mind.

Inanimate control sequences

A natural first step towards a useful definition would be to require domain-specificity; to describe mechanisms as constituting an implicit theory of mind only if there is evidence that they are dedicated to, or especially good at, processing the behaviour of other agents. On this moderately liberal definition, infant FB studies would provide evidence that infants have an implicit theory of mind if they could not be explained without reference to what are sometimes called 'behaviour rules'; mentally represented principles that specify what an agent will do under specified, observable conditions. In his penetrating discussion of the relationship between mentalistic and behaviour rules, Perner (2010, p. 246) gives the following example of a behaviour rule: 'If a person P looks at an object O being put inside a location L1, and does not look when it is transferred to location L2, and if P is to get the object O, then P will go to L1' (Perner, 2010, p. 246). The low-level novelty hypothesis suggests that the results to date of infant FB studies *can* be explained without recourse to behaviour rules, or any other domain-specific psychological processes, and therefore, even relative to the moderately liberal definition, these results do not provide evidence of implicit theory of mind.

An obvious way to test the low-level account against the hypothesis that infants apply behaviour rules, or some other domain-specific processes, would be to use inanimate controls. In these control conditions, infants would see sequences of events that have all the low-level features of sequences in which agents act on objects, except those features that are characteristic of biological agents. This approach has been used successfully in adults to distinguish a domain-specific theory of imitation from a domain-general account (Leighton, Bird & Heyes, 2010), but, as far as I am aware, it has not been used in research on FB in infancy.

Surian and Geraci (2012) moved in this methodological direction when they tested infants for the attribution of FB to a red triangle interacting with a blue disc, and failed to find evidence of domain-specificity. However, they did not explicitly compare the behaviour of infants presented with matched animate and inanimate sequences, and they removed morphological but not dynamic cues characteristic of biological agents.[13] A full and more informative approach would do both of these things. Explicit comparison of matched animate and inanimate sequences – between-subjects or, ideally, within-subjects – would give us an opportunity to find out whether the mechanisms controlling performance in infant FB studies are especially good at, rather than exclusively dedicated to, the processing of behavioural stimuli. The manipulation of dynamic as well as morphological cues would allow for the possibility that looking behaviour is controlled by domain-specific mechanisms that are triggered by the former rather than the latter.[15]

The use of inanimate control sequences would require that all stimuli are presented on video, but, as the studies by Surian *et al*. (2007), Kovács *et al*. (2010) and others have demonstrated, the use of video stimuli is eminently feasible in this context. It would also be crucial to ensure, through careful piloting of the inanimate sequences and monitoring throughout the experiments, that infants attend equally to animate and inanimate sequences. This is certainly possible using eye-tracking technology, but it would necessitate more detailed recording and reporting of eye movements than is currently typical.

The utility of inanimate controls does not depend on the assumption that, if infants are capable of ascribing mental states, they will ascribe mental states always and only to objects that are understood by the adults of their culture, or within a scientific world view, to be agents. It is possible that what counts as an agent for an infant overlaps with, but is not identical to, what counts as an agent for an adult. Therefore, an 'inanimate' control could involve any object for which there are principled reasons to believe that, even if infants can ascribe mental states, they will be disinclined to ascribe them to this object. The precise location of the animate/inanimate boundary can be adjusted, but if implicit mentalizing hypotheses do not assume a negative class – that there are objects to which infants seldom or never ascribe mental states – there is a danger that implicit mentalizing hypotheses will become untestable (Santiesteban, Catmur, Bird & Heyes, under review).

Self-informed belief induction

If the use of inanimate control sequences, or other experimental techniques, provides clear evidence of domain-specificity, according to the moderately liberal definition we would have evidence that infants possess implicit theory of mind. However, my sense is that many researchers would like to go further in restricting the definition of what constitutes implicit theory of mind – to use a tight definition excluding behaviour rules – and want to know whether infant looking behaviour in FB experiments is controlled by behaviour rules or whether it 'really' involves the attribution of mental states. It is notoriously difficult to specify exactly what it would be 'really' to attribute mental states, and how this differs from the use of behaviour rules. However, a common intuition is that real mentalizing involves going beyond observable relations between circumstances and behaviour, and representing these regularities as indicative of unobservable (theoretical or subjective) causal states with intentional content; states such as beliefs and desires. How might mentalizing in this sense be distinguished empirically from the use of behavioural rules alone, in tests that do not require language?

I'd like to recommend a strategy that, following Novey (1975), I proposed initially as a way of testing for theory of mind in nonhuman primates (Heyes, 1998). The strategy involves using a 'self-informed' belief induction variable; a variable that, if the infant is capable of mentalizing, she knows only through extrapolation from her own experience to be indicative of what an agent can or cannot see, and therefore does or does not believe. The logic of self-informed belief induction could be implemented in many ways, using various concrete obstacles to perception (goggles, visors, headphones, gloves) in different modalities (vision, hearing, touch), and it could be combined with any of the procedures currently used to test for FB attribution in infancy. However, for the purpose of illustration, I'll use visors, vision, and Onishi and Baillargeon's (2005) original nonverbal test for FB attribution.

Before the experimental procedure, the infants would be given first-person experience with two visors, one translucent and the other opaque. (I have in mind the kind of visors that are attached to motorcycle helmets. They would be lightweight, of a size that does not cover the whole face, and attached to the head at a comfortable distance from the face.) The translucent and opaque visors would be of different colours, say red and blue (counterbalanced), but otherwise they would be identical in appearance when worn by another agent. Importantly, however, during the pre-experimental phase, the infants would not see the visors worn by another agent, or have any opportunity to discover that agents behave differently when the red and blue visors are interposed between the agents and other objects. Instead, the infants would wear each of the visors themselves in the context of a game. Through this experience they would have the opportunity to learn that the red visor affords seeing, and the blue visor does not. Following the period of pretraining in which infants receive this experience, they would undergo the familiarization, belief induction and test procedures used in Onishi and Baillargeon's original experiment (see Figure 1), but the red/blue visors, rather than presence/absence of the agent, would be used as the belief induction variable.[16]

Like the belief induction variables typically used in infant FB experiments, a self-informed belief induction variable, such as red visor/blue visor, reliably correlates with whether an agent can see, and therefore has true beliefs about, key events. However, unlike the belief induction variables in common use, if the logic of a self-informed belief induction variable is properly implemented, its significance with respect to seeing and believing (1) is not confounded by low-level features that vary in the degree to which they cause retroactive interference or distraction, and (2) cannot readily be encoded in a purely behavioural rule.

My original recommendations regarding self-informed belief induction (Heyes, 1998) have been pursued in research involving chimpanzees (Penn & Povinelli, 2007;

Povinelli & Vonk, 2004), human adults (Langton, 2009; Teufel, Alexis, Clayton & Davis, 2010) and infants (Meltzoff & Brooks, 2008; Senju, Southgate, Snape, Leonard & Csibra, 2011). A recent study by Senju *et al.* (2011) is particularly interesting because it used a self-informed belief induction variable to test, not merely for the attribution of 'seeing', but for the attribution of FB in infants. The 18-month-olds in this study were tested in an anticipatory looking procedure similar to the one used by Southgate *et al.* (2007; see 'Bell ringing' section above). Before the main procedure began, they received pretraining in which 'opaque' or 'trick' blindfolds (between-subjects) were interposed between the infants' eyes and a variety of objects, and the experimenter asked about the location of the object. For example, when the blindfold was between the infant and a toy bunny, the experimenter asked 'Where's the bunny?'[17] The opaque blindfolds were truly opaque, but the trick blindfolds, lacking a thick interior layer of fabric, were translucent. Both types of blindfold were black with a frilly pink border.

In the main procedure, a human agent was seated behind a panel containing two windows, on the infant's left and right, and below each window was an opaque box. The agent's head was visible at the top of the panel. At the beginning of each familiarization trial, a puppet placed a toy in one of the two boxes. When the puppet had disappeared, both windows were illuminated, a tone sounded and after a fixed interval a hand came through the window above the box containing the toy, moved towards that box and retrieved the toy. In the belief induction trial, once the puppet had placed the toy in the left box, instead of looking away (Southgate *et al.*, 2007), the agent put on a black blindfold with a pink frilly border, and this blindfold remained in place as the puppet retrieved the toy from the left box and removed it from the scene. Once the puppet had disappeared, the agent removed the blindfold and the test trial started, i.e. the windows were illuminated, and the tone sounded. The results showed that the infants who had been pretrained with the opaque blindfolds made their first eye movement towards the left box, and looked longer at the left than the right box. In contrast, infants who had experienced the trick blindfolds were equally likely to make their first eye movement to the left and the right boxes, and did not look longer at the left than the right box.

This implementation of the logic of self-informed belief induction is very promising indeed. It shows that, with care and ingenuity, it is possible to design obstacles to vision and a pretraining regime with roughly the right characteristics to implement the logic. Furthermore, this experiment demonstrates that, even when there is a significant interval between pretraining and belief induction, pretraining experience with visual obstacles can modulate the effects of belief induction on looking direction and duration in infants. However, it is not a case of 'job done'; there is an intriguing ambiguity in the results reported by Senju *et al.* (2011), and certain features of their procedure are useful in highlighting potential problems in implementing the logic of self-informed belief induction. The ambiguity (which was acknowledged by Senju *et al.*) relates to the behaviour of the infants who experienced the trick blindfold. Their failure to discriminate between the two boxes on test could be interpreted as evidence that the trick group attributed a true belief (that the toy was not in either box) to the agent. This is certainly what one would expect if the opaque group showed a preference for the left box because they attributed a FB to the agent. However, another possibility is that the opaque group did, and the trick group did not, discriminate between the boxes on test because the infants in the trick group were less distracted by the blindfold, and therefore more likely themselves to see removal of the toy from the scene during the belief induction trial. This possibility highlights the need to check that the occluding and non-occluding objects used during pretraining have their effects by virtue of their specific, perception-related properties, and not by virtue of being more or less interesting or aversive to the infants (Heyes, 1998).

Certain features of the pretraining procedure used by Senju *et al.* (2011), which was modelled on that of Meltzoff and Brooks (2008), highlight another pair of important implementation issues. First, infants should not be able to perceive during pretraining differences between the occluding and non-occluding objects that are likely to have been correlated in the infants' everyday experience with different behaviours by an agent. Second, the agent/experimenter must not behave differently in relation to the occluding and non-occluding objects during pretraining. If these conditions are not met, there is a risk that infants will 'pass' the FB test, not by attributing true and FBs, but by applying behaviour rules (Perner, 2012). It is possible that these conditions were not met in the pretraining procedure used by Senju *et al.* (2011) and Meltzoff and Brooks (2008) because it involved sheer and opaque fabrics, which are common in most households, in combination with 'interposition'. Rather than being close to the infants' faces throughout the pretraining procedure, the 'blindfolds' were often located at some distance from the infants' and from the agent's bodies, where they could be viewed from the same perspective as sheer and opaque fabrics are seen in everyday life, and in a three-term relationship with the agent and an object. This is why, although it places significant demands on researchers, I have recommended that infants *wear* the occluding and non-occluding objects for self-informed belief induction, and that they are discriminable only on the basis of an arbitrary cue.

The infant research reviewed in this article already constitutes a dazzling display of ingenuity. It takes years of experience, careful reasoning, and a great deal of sensitivity and imagination to design experiments that coax infant minds to give up *any* of their secrets, let alone secrets about whether and how infant minds represent the minds of others. I have argued that, in spite of the elegance and ingenuity of many of the experiments to date, we still do not know whether infants can attribute FB, and recommended some new strategies. I hope these will prove useful as the field moves closer to answering one of the most intriguing and theoretically important questions currently addressed by developmental science.

Acknowledgements

I am grateful to Martin Eimer, Chris Frith, Uta Frith, Jennifer Nagel and Nick Shea for valuable discussion while I was writing this paper, and to three anonymous reviewers for their comments and helpful challenges.

References

Apperly, I.A., & Butterfill, S.A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, **116**(4), 953–970.

Baillargeon, R. (1999). Young infants' expectations about hidden objects: a reply to three challenges. *Developmental Science*, **2**(2), 115–132.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, **112**(2), 337–342.

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language*, **28**(2), 141–172.

Clements, W.A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, **9**(4), 377–395.

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**(1567), 1149–1157.

Daum, M.M., Attig, M., Gunawan, R., Prinz, W., & Gredebäck, G. (2012). Actions seen through babies' eyes: a dissociation between looking time and predictive gaze. *Frontiers in Psychology*, 3.

De Bruin, L.C., & Newen, A. (2012). An association account of false belief understanding. *Cognition*, **123**(2), 240–259.

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development*, **56**(4), 868–883.

Diamond, A. (1990). The development and neural bases of memory functions as indexed by the AB and delayed response tasks in human infants and infant monkeys. *Annals of the New York Academy of Sciences*, **608**(1), 267–317.

Fitzpatrick, S. (2008). Doing away with Morgan's Canon. *Mind & Language*, **23**(2), 224–246.

Fuster, J.M., & Bauer, R.H. (1974). Visual short-term memory deficit from hypothermia of frontal cortex. *Brain Research*, **81**(3), 393–400.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, **56**(2), 165–193.

Grant, D.S. (1988). Sources of visual interference in delayed matching-to-sample with pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**(4), 368–375.

Haith, M.M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development*, **21**(2), 167–179.

Heyes, C. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, **21**(1), 101–114.

Heyes, C. (2012). Simple minds: a qualified defence of associative learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**(1603), 2695–2703.

Kovács, Á.M., Téglás, E., & Endress, A.D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science*, **330**(6012), 1830–1834.

Langton, S.R.H. (2009). Gaze perception: is seeing influenced by believing? *Current Biology*, **19**(15), R651–R653.

Leighton, J., Bird, G., & Heyes, C. (2010). Goals are not an integral component of imitation. *Cognition*, **114**(3), 423–435.

Longo, M.R., & Bertenthal, B.I. (2006). Common coding of observation and execution of action in 9-month-old infants. *Infancy*, **10**(1), 43–59.

Luo, Y. (2011a). Do 10-month-old infants understand others' false beliefs? *Cognition*, **121**(3), 289–298.

Luo, Y. (2011b). Three-month-old infants attribute goals to a non-human agent. *Developmental Science*, **14**(2), 453–460.

Marcovitch, S., & Zelazo, P.D. (2003). The A-not-B error: results from a logistic meta-analysis. *Child Development*, **70**(6), 1297–1313.

Meltzoff, A.N., & Brooks, R. (2008). Self-experience as a mechanism for learning about others: a training study in social cognition. *Developmental Psychology*, **44**(5), 1257–1265.

Meristo, M., Morgan, G., Geraci, A., Iozzi, L., Hjelmquist, E., Surian, L., & Siegal, M. (2012). Belief attribution in deaf and hearing infants. *Developmental Science*, **15**(5), 633–640.

Miller, E.K., Erickson, C.A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, **16**(16), 5154–5167.

Novey, M.S. (1975). The development of knowledge of others' ability to see. Unpublished doctoral dissertation, Harvard University.

Olesen, P.J., Macoveanu, J., Tegnér, J., & Klingberg, T. (2007). Brain activity related to working memory and distraction in children and adults. *Cerebral Cortex*, **17**(5), 1047–1054.

Onishi, K.H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, **308**(5719), 255–258.

Paulus, M., Hunnius, S., van Wijngaarden, C., Vrins, S., van Rooij, I., & Bekkering, H. (2011a). The role of frequency information and teleological reasoning in infants' and adults' action prediction. *Developmental Psychology*, **47**(4), 976–983.

Paulus, M., Hunnius, S., Vissers, M., & Bekkering, H. (2011b). Imitation in infancy: rational or motor resonance? *Child Development*, **82**(4), 1047–1057.

Pearce, J.M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, **94**(1), 61–73.

Pearce, J.M. (2008). *Animal learning and cognition: An introduction*. New York: Psychology Press.

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: Biological Sciences*, **362**(1480), 731–744.

Perner, J. (2010). Who took the cog out of cognitive science? *International Perspectives on Psychological Science*, **1**, 241–262.

Perner, J. (2012). MiniMeta: in search of minimal criteria for metacognition. In M. Beran, J.L. Brandl, J. Perner & J. Proust (Eds.), *Foundations of metacognition* (pp. 94–116). Oxford: Oxford University Press.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: how deep? *Science*, **308**(5719), 214–216.

Poulin-Dubois, D., & Chow, V. (2009). The effect of a looker's past reliability on infants' reasoning about beliefs. *Developmental Psychology*, **45**(6), 1576–1582.

Povinelli, D.J., & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind & Language*, **19**(1), 1–28.

Rankin, C.H., Abrams, T., Barry, R.J., Bhatnagar, S., Clayton, D., Colombo, J., Coppola, G., Geyer, M.A., Glanzman, D.L., Marsland, S., McSweeney, F., Wilson, D.A., Wu, C.-F., & Thompson, R.F. (2009). Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory*, **92**(2), 135–138.

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, **30**(1), 87–104.

Ruz, M., & Lupiáñez Castillo, J. (2002). A review of attentional capture: on its automaticity and sensitivity to endogenous control. *Psicológica: Revista de metodología y psicología experimental*, **23** (2), 283–310.

Santiesteban, I., Catmur, C., Hopkins, C.S., Bird, G. & Heyes, C. (in press) Avatars and arrows: implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, doi: 10.1037/a0035175.

Scott, R.M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, **80**(4), 1172–1196.

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), 366–395.

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? *Psychological Science*, **22**(7), 878–880.

Sirois, S., & Jackson, I. (2007). Social cognition in infancy: a critical review of research on higher order abilities. *European Journal of Developmental Psychology*, **4**(1), 46–64.

Slaughter, V., Heron-Delaney, M., Christie, T., Slaughter, V., & Brownell, C. (2011). Developing expertise in human body perception. In V. Slaughter & C. Brownell (Eds.), *Early development of body representations* (pp. 81–100). Cambridge: Cambridge University Press.

Song, H., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology*, **44**(6), 1789–1795.

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**(3), 295–315.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, **13**(6), 907–912.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, **18**(7), 587–592.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, **18**(7), 580–586.

Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology*, **30** (1), 30–44.

Teufel, C., Alexis, D.M., Clayton, N.S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception, & Psychophysics*, **72**(3), 695–705.

Träuble, B., Marinović, V., & Pauen, S. (2010). Early theory of mind competencies: do infants understand others' beliefs? *Infancy*, **15**(4), 434–444.

Woodward, A.L. (2009). Infants' grasp of others' intentions. *Current Directions in Psychological Science*, **18**(1), 53–57.

Yoon, J.H., Curtis, C.E., & D'Esposito, M. (2006). Differential effects of distraction during working memory on delay-period activity in the prefrontal cortex and the visual association cortex. *NeuroImage*, **29**(4), 1117–1126.

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), 156–171.

Received: 31 October 2012
Accepted: 29 September 2013

Supporting Information

The superscript numbers in this article refer to notes that are listed in the online version of this article:

Data S1. False Belief in Infancy: A Fresh Look