A puzzle about belief-ascription[1]

Pierre Jacob

By the end of the twentieth century, following the comments by three philosophers on Premack and Woodruff's seminal (1978) paper entitled "Does the chimpanzee have a theory of mind",[2] twenty years or so of careful developmental psychology were taken to show that not until they are well in their fourth year can human children successfully pass the so-called "standard false belief task". But in 2005, Onishi and Baillargeon published in *Science* the first of a series of papers reporting findings based on the violation-of-expectation paradigm that strongly suggest that preverbal human infants can represent another's false beliefs. The puzzle is: how come 3-year-olds fail the standard false belief task if preverbal human infants are able to represent another's false belief?

As Perner and Ruffman (2005) have put it in their critical comments on Onishi and Baillargeon's (2005) paper, "understanding of false belief at 4 years of age suggests that this

[1] Versions of this paper were presented at the Philosophy Colloquium in Stockholm University on May 19 2011 and at a Workshop on early mindreading at the Ecole Normale Supérieure in Paris on December 9 2011. I am grateful to Kathrin Glüer, Åsa Wikforss, Peter Pagin, Gergo Csibra, Gyuri Gergely, Agnes Kovacs, Dan Sperber, Frédérique de Vignemont and Ian Apperly for their critical reactions. I am also most grateful to Kris Onishi for numerous discussions and for her critical comments on an earlier draft of this paper. Above all, I am grateful to Steve Butterfill for his detailed and valuable written responses to an earlier draft of this paper, which forced me to rewrite substantially section 3. I am sure, however, that he still disagrees with what I say here.

[2] Cf. Bennett (1978), Dennett (1978) and Harman (1978).

ability may be constructed in a cultural process tied to language acquisition. In contrast, competence at 15 months suggests that this ability is part of our purely biological inheritance". There are presently three main strategies to deal with this puzzle, the third of which has only recently emerged. The first nativist strategy takes the experiments in the violation-of-expectation paradigm with preverbal human infants at face value and offers an explanation of why significantly older children find it so difficult to pass the standard elicited false belief task. The empiricist second strategy posits that not until children can pass the standard false belief task can they represent another's false belief and it offers some deflationary low-level account of the results of the violation-of-expectation experiments with preverbal human infants. Very recently, a third strategy has emerged: a two-systems model of belief-ascription. The goal of this paper is to assess these three strategies in light of a selected sample of the relevant experimental findings.[3]

1. The basic empirical challenge from the standpoint of the first strategy

1.1. Why the standard elicited false belief task?

By asking the question whether chimpanzees have a *theory of mind*, Premack and Woodruff (1978) raised the issue whether they can compute and represent others' psychological states (e.g. intentions, desires, beliefs), ascribe them to others, make inferences about them and thereby explain others' behavior.[4] To *ascribe* a psychological state to another, e.g. an intention to grab a banana, is to *believe* or *judge* that another intends to grab a banana.[5] Arguably, a creature can intend to grab a banana without possessing the concept *intention*, but one could not judge or believe that another intends to grab a banana (i.e. ascribe this intention to an agent) unless one possessed the concept *intention*.

[3] For an extensive survey of the experimental findings, cf. Caron (2009).
[4] Another current term for what Premack and Woodruff (1978) called "theory of mind" is "mindreading". I will use both terms interchangeably.
[5] Having a theory of mind may also involve the ability to ascribe psychological states to oneself or engage in tasks of first-person mindreading. But I will not consider this question here at all.

Now, in many cases, one can *predict* an agent's action by *sharing* her motivation (goal and intention) and her epistemic state (knowledge about the world). To share another's goal or intention is to have a goal or intention and to have a goal or intention is to represent a possible non-actual state of affairs that will be turned into a fact or an actual state of affairs by one's action (cf. Searle, 1983; Jacob and Jeannerod, 2003; Jacob, in press). To share another's epistemic state of e.g. knowledge is to represent some actual state of affairs (or fact) relevant to the satisfaction of one's goal or intention. One can often successfully predict an agent's action by representing the possible state of affairs that constitutes the content of the agent's goal or intention and the relevant facts that are known by the agent. If so, then one can often successfully predict an agent's action by representing possible and actual aspects of the world (which are also being represented by the agent), without (meta-) representing the agent's relevant psychological states and ascribing them to him or her. But then predicting an agent's action falls short of *explaining* the agent's action, which requires representing the right motivational and epistemic psychological states and ascribing them to the agent, whether or not one shares the agent's motivational and epistemic psychological states.[6]

What would decisively show that a creature has a theory of mind in the required sense is that the creature could correctly compute, represent and ascribe to another some motivational or epistemic psychological state that she does *not share*. For example, it would be sufficient if one could provide evidence that a creature can ascribe to another a preference for one of a pair of objects different from her own. Alternatively, one could investigate a creature's ability to ascribe to another an epistemic psychological state different from her own about a given situation. This is exactly what is accomplished by the standard "false belief task", in which a participant is provided with three pieces of information: (i) she knows from

[6] I take it that an advocate of a sophisticated version of a simulation-based approach to mindreading, such as Goldman (2006), would agree with me. This is why Goldman's (2006) full approach is in fact a simulation-and-projection approach, whereby the mindreader could not complete her mindreading task unless she projected onto her target the psychological state generated by the simulation routine.

observing her that a character (call her Sally) is *motivated* to find her marble. (ii) She knows in which of two possible locations the marble really is. (iii) Finally, she knows or believes that Sally falsely believes her marble to be in the location where it is not because she saw the marble being moved from one location to the other in Sally's absence (cf. Wimmer and Perner, 1983; Baron-Cohen, Leslie and Frith, 1985). The participant's task is to predict where the agent with a false belief about the object's location will look for the object.

Following Baillargeon et al. (2010), I shall call this task the *elicited* false belief task because the participant is expected to answer an *explicit* question such as: "where will the agent look for the object?" When asked to predict where an agent with a false belief about an object's location will look for it, children who know the object's location have been demonstrated to reliably fail until they are in their fourth year. Furthermore, despite a mental age well in excess of four years, autistic children have been repeatedly reported to fail on the elicited false belief task, whereas other disabled children, for example with Down syndrome, could succeed (cf. Leslie, 2000 for a survey).

1.2. Spontaneous responses in the violation-of-expectation paradigm

Clearly, passing the standard false belief task (succinctly described above) is a *sufficient* condition for being credited with a theory of mind in Premack and Woodruff's (1978) sense. However, it might not be a *necessary* condition. A creature with the ability to represent and ascribe to another a false belief might still fail the standard false belief task, because the four following conditions are further required in order for a participant to pass the standard false belief task (emphasized by Baillargeon et al., 2010; Bloom and German, 2000; Leslie, 2000; Leslie, 2005):

Condition 1. The participant must speak the natural language in which the relevant test question is being asked.

Condition 2. The participant must have the pragmatic capacity to understand that the relevant question being asked is: where *will* Sally look for the object, not where *should* Sally look for the object, nor where *is* the object?

Condition 3. The participant must have enough executive control in order to inhibit the prepotent tendency to answer the relevant question on the basis of her own true belief about the object's location.

Condition 4. The participant must combine the representation of the agent's goal (to find the object) with the representation of the agent's false belief about its current location.

For many years, evidence that young children, who fail the standard false belief task, might nonetheless represent (or understand) the contents of others' false beliefs was reported by psychologists. For example, Clements and Perner (1994) noted that children aged 2 years and 11 months, who failed the elicited standard false belief task, *looked* at the empty location where the agent falsely believed the object to be located. This suggests that they might have correctly represented the agent's false belief, but lacked further resources required for correctly answering the explicit question. Also, Surian and Leslie (1999) report that, in the elicited false belief task, when 3-year-olds are asked the question "where will the agent *first look*?" (instead of "where does the agent *think* the object is?"), their performance (unlike the performance of autistic children) significantly improved.

Some twenty odd years after the publication of the first experiments using the standard false belief task, Onishi and Baillargeon (2005) managed to adapt the violation-of-expectation paradigm to design a *spontaneous* false belief task. According to the basic assumption of the violation-of-expectation framework (well exemplified by magic tricks in human adults), when an individual's expectations are violated, she is surprised and, as a result, she looks longer at an unexpected than at an expected event. Violation-of-expectation experiments have two steps: first, in the habituation or familiarization trials, infants are experimentally induced to

form expectations by being repeatedly exposed to one and the same event.[7] Secondly, in the test trials, they are presented with either an expected or an unexpected ('magical') event. By measuring the time during which infants respectively look at the expected vs. the unexpected event, psychologists get evidence about the nature and content of the infants' expectations formed during the habituation or familiarization trials.

Violation-of-expectation experiments are well suited for testing the ability of preverbal human infants to represent others' false beliefs for two reasons. First of all, in this framework, experiments can be run without asking participants any explicit question: as Baillargeon et al. (2010, p. 110) have emphasized, infants' understanding of an agent's false belief can be inferred from their spontaneous behavior, "as they observe a scene unfold (just as adults watching a movie might spontaneously produce responses that reveal their understanding of the characters' mental states)". Consequently, such experiments can test participants' ability to represent and understand others' false beliefs independently of the further cognitive resources (Conditions 1-3) required for passing the standard elicited false belief task. Secondly, experiments in this framework test infants' *expectations* about an agent's action and/or psychological states. An infant's expectation about an agent's psychological state has the same mind-to-world direction of fit as beliefs about others' psychological states involved in standard tasks of mindreading (not the world-to-mind direction of fit of motivational states): it can be correct or incorrect.

Several years before Onishi and Baillargeon designed a false belief scenario within the violation-of-expectation paradigm, Woodward (1998, 1999) used this paradigm to investigate the ability of preverbal infants to understand an agent's goal and ascribe preferences to an agent. In the habituation trials of her study, 6-month-olds saw a human hand in a grasping posture repeatedly select a teddy bear on the right, in the presence of a ball on the left. In the

[7] Habituation trials are repeated until the infant satisfies some preset criterion, defined in terms of looking times. Familiarization trials are repeated during a fixed number of trials.

test trials, the locations of the toys were switched and the infants saw the hand select either the same toy (the teddy bear) at a new location (on the left) or a new toy (the ball) at the old location (on the right). Woodward (1998, 1999) found that 6-month-olds were more surprised to see (i.e. they looked longer when they saw) the human hand select the new toy at the previous location than the same toy at a new location in the test trial. Woodward (1998) argued that this finding shows that 6-month-olds interpret a human hand's preference for one of two targets as evidence that the human hand's action is goal-directed.[8]

A few years later, Luo and Baillargeon (2005) replaced the human hand by a simple box in Woodward's design: in the familiarization trials, 5-month-olds saw a box repeatedly contact a cone on its right in the presence of a cylinder on its left. In the test trials, after the cylinder and the cone had switched their locations, 5-month-olds looked longer when the box moved to the cylinder at the old location rather than to the cone at a new location. Luo and Baillargeon interpreted their findings as showing that 5-month-olds interpret a preference displayed by a box for one of two targets as evidence that the box's motion is goal-directed.

In the familiarization trials of Onishi and Baillargeon's (2005) experiment, 15-month-olds saw a human female agent grasp and briefly play with a toy located in between a yellow and a green box before hiding it inside the green box and pause with her hand inside the green box. In the second and third familiarization trials, the infants saw the agent reach inside the green box (as if she was retrieving the toy) and then pause. Next, each infant was exposed to one of four different belief-induction trials, in which they saw the toy move from one box to the other in either the presence or the absence of the agent. Finally, during two test trials, infants saw the agent reach for the box in either the yellow or the green box on the basis of either a true or a false belief. Onishi and Baillargeon (2005) report that infants looked reliably longer when they saw the agent reach for the toy either in the wrong location while she had a

[8] Woodward (1998, 1999) reports that infants did not look longer if, instead of a human hand, either a rigid rod or a back-of-the-hand selected a novel toy at the old location rather than the same toy at a new location. For further discussion, cf. Jacob (in press).

true belief about the toy's location or in the right location while she had a false belief. Surian et al. (2007) have further reported that 13-month-olds look longer at test trials in which an agent retrieves its preferred food when it is hidden from the agent's (but not the infant's) view by a high barrier than when it is visible to the agent and they also look longer when the food hidden from the agent's view by a barrier has been placed there in the agent's absence than in the agent's presence.

Buttelman et al. (2009) seem to think that findings based on the violation-of-expectation paradigm alone could not constitute decisive evidence that preverbal human infants can ascribe false beliefs to an agent and should, therefore, be supplemented by other non-verbal more active behavioral measures. In their own experiments, children watched a toy being switched from one box to another either in the presence (true belief condition) or the absence (false belief condition) of an adult. Then the adult attempted unsuccessfully to open the box in which the toy had been originally located. They based their experiments on children's early propensity to help another achieve her goal (investigated by Warneken and Tomasello, 2006, 2007). If children represent another's false belief, then the way their propensity to help is put to use should differ in the true belief and the false belief conditions. In the former, children should assume that the agent with a true belief is not trying to retrieve the toy and if so, then they should help the agent open the empty box. But in the latter condition, children should assume that the agent with a false belief is trying to retrieve the toy and if so, then they should help the agent by opening the other box (with the toy) and retrieve the toy from it. Buttelman et al. (2009) found that the helping behavior of 2,5 month-olds and 18-month-olds reliably differed in the true and the false belief conditions, but in 16-month-olds the difference failed to reach significance against chance level in the true belief condition.

Beyond the methodological differences between the helping and the violation-of-expectation paradigms, the findings reported by Buttelman et al. (2009) and by Onishi and Baillargeon (2005) both strongly suggest that before they can produce complex sentences expressing ascriptions of psychological states, human infants make sense of an agent's behavior by tracking jointly her goal (or motivation) and her epistemic state, even when the latter happens to be incongruent with the state of affairs that it represents.

2. Assessing the second strategy

Advocates of the empiricist second strategy take success at the false belief task as a criterion of the ability to ascribe false beliefs to others. Thus, they find it incredible that preverbal human infants, who lack most, if not all, of the resources mentioned in Conditions 1-3 of section 1.2, could nonetheless represent others' false beliefs (cf. Perner and Ruffman, 2005; Ruffman and Perner, 2005). So on the one hand, their burden is to offer low-level accounts, which can explain the findings about preverbal human infants without crediting them with the ability to ascribe false beliefs to others. On the other hand, they do not seem to find it incredible that preverbal human infants might be able to represent others' motivational states. If not, why not?

2.1. Three low-level explanations of data from preverbal human infants

Before Woodward's (1998) paper on goal-ascription, Gergely et al. (1995) and Csibra et al. (1999) used a different design to investigate infants' ability to represent an agent's goal-directed action. In the habituation trials of one of their classic studies, infants in their first year of life saw a small circle repeatedly move from right to left in a parabolic trajectory above a rectangle before contacting a large circle. In the test condition, the rectangle was removed and infants either saw the small circle repeat the same parabolic trajectory or go in a straight line

before contacting the large circle. Infants looked longer at the old parabolic trajectory than at the novel rectilinear trajectory. Infants who saw the same display in the habituation trials except that the rectangle stood behind the small circle, not in between the two circles, looked equally at the two test conditions. Gergely and Csibra (2003) have argued that before the end of their first year, preverbal human infants automatically adopt the "teleological stance" that enables them to interpret an agent's action (e.g. the parabolic trajectory of the small circle) as an efficient means to achieving a goal (e.g. contacting the large circle), relative to a set of situational constraints (e.g. the presence of a rectangle). They have offered further evidence that given any two of the three parameters (action-means, goal state and constraint), infants can successfully infer the third.

Advocates of the second strategy might want to endorse Gergely and Csibra's (2003) further distinction between a *teleological* and a *mentalistic* representation of an action. On the mentalistic interpretation, the small circle's action would be seen as intentional: the agent selects (or intends) its parabolic course of action as an efficient means to fulfill its *desire* to contact the large circle, given its *belief* that there is a rectangle standing in its way. On the mentalistic interpretation, some version of the rationality principle applies to the contents of the agent's belief and desire. On the teleological interpretation, some version of the efficiency principle applies to an action as a means to achieve a goal state. But it is not entirely clear how to formulate a teleological interpretation entirely devoid of any mentalistic component. The agent's action may be construed as an efficient means to achieve a goal state, relative to the constraining presence of the rectangle, but only relative to the agent's goal (or intention) of achieving a goal state. In Gergely and Csibra's (2003, p. 289) terms, the teleological stance is supposed to establish "explanatory relation among three relevant aspects of current and future reality: the action, the (future) goal state, and the current situational constraints". If so, then given the infant's perception of the agent's *action* (whereby the agent repeatedly

achieves the same goal state), two questions jointly arise: one is whether the infant need not be credited with the ability to ascribe to the agent the mentalistic intention to persist in its attempt at achieving the same goal state. The other is whether the infant should be credited with the ability to *share* the agent's epistemic information about there being a situational constraint and the agent's ability to compute an efficient means for achieving the agent's goal state.

2.1.1. The three-way association hypothesis

In a similar line, Perner and Ruffman (2005) have offered two low-level explanations of the findings reported by Onishi and Baillargeon (2005), the first of which is the *three-way association* between an agent, an object and a location. Perner and Ruffman (2005) hypothesize that when in the familiarization trials infants see an agent hide a toy in one of two locations, they form an association linking the agent, the toy, and its hiding location. They further hypothesize that this association causes the infants to look longer to the test event that most deviates from the association (e.g. when the agent on the Onishi and Baillargeon's scenario searches for the toy in a different location). If so, then it would seem as if there would be no need to credit infants with the ability to ascribe either true or false beliefs to the agent. Perner and Ruffman's (2005) three-way association hypothesis, however, faces three related problems.

The first problem is that, while in the first familiarization trial the object was visible before it was hidden into one of the two boxes, in the test trials the object is no longer visible to either the agent or the infant before the agent retrieves it from out of the box where it is hidden. So for the three-way association to be sustained in the test trials, the infant must assume that the agent *knows* and *remembers* the object-box relation that was instantiated in the familiarization trial. In other words, the association hypothesis itself presupposes that the infant must ascribe an epistemic state of knowledge to the agent.

The second problem arises from Woodward's (1998) findings described above and involving *two* objects: 6-month-olds were more surprised to see the human hand select the new toy at the previous location than the same toy at a new location in the test trial. Advocates of the first strategy interpret this finding as evidence that infants ascribe a preference to the agent. Clearly, the three-way association hypothesis should also apply to this case. But the problem for the three-way association hypothesis is to determine *which* of a change of toy or a change of location deviates most from a three-way association agent-toy-location. How can one tell?

The third problem arises from the findings by Luo and Baillargeon (2005): 5-month-olds look longer when the box moved to a new target (the cylinder) at the old location rather than when the box moved to the old target (the cone) at a new location. But Luo and Baillargeon (2005) also tested another condition — the single-object condition —, in the familiarization trials of which infants saw the box move repeatedly towards the cone on the right, but now the cylinder was missing. Luo and Baillargeon (2005) found that in the test trials, infants looked equally when the box moved either to the cone at a new location or to the cylinder at the old location. Arguably, in the single-object condition, infants lacked evidence for ascribing a preference to the box. But given that infants formed the very same three-way association (agent-object-location) in the two-objects familiarization trials and in the single-object familiarization trials, the three-way association hypothesis should predict that infants should respond the same way to the test trials in both conditions. But they do not.

2.1.2. The search heuristic

Perner and Ruffman (2005, p. 215) have also offered a second low-level account of the findings reported by Onishi and Baillargeon (2005) under the guise of the following search heuristic (or behavior rule) according to which "infants may have noticed (or are innately predisposed to assume) that people look for an object where they last saw it and not

necessarily where the object actually is". But as noticed by Surian et al. (2007, p. 585), the search heuristic hypothesis faces the following dilemma: either infants have innate knowledge of this rule or they learnt it through experience. If the former, then the hypothesis would be committed to an arbitrarily strong form of nativism. If the latter, in the light of the fact that people commonly pick up objects in plain sight or look for them in several different places before finding them, then the question arises: what could be the evidential basis that could enable infants to learn the rule?

2.1.3. Ignorance ascription

Unlike an agent's knowledge, an agent's false belief is not congruent with a fact. Hogrefe et al. (1986) reported evidence that at 3-4 year-olds were able to attribute ignorance but failed to attribute a false belief to an agent. As emphasized by Baillargeon et al. (2010), it may seem easier to represent the *incomplete* content of another's *ignorance* of a fact by subtracting or bracketing aspects of one's own representation than the *false* content of another's representation, which clashes with the content of one's own. Thus, a third deflationary interpretation of Onishi and Baillargeon's (2005) findings has been suggested by Southgate et al. (2007, p. 587). While they provide evidence that two-year-olds correctly anticipate an actor's actions when these actions can be predicted only by attributing a false belief to the actor, Southgate et al. (2007) hypothesize that, instead of ascribing false beliefs to the agent, infants in Onishi and Baillargeon's (2005) experiments might merely attribute ignorance to the agent and further assume, in accordance with evidence reported by Ruffman (1996), that ignorant agents are likely to search in the wrong location rather than perform at chance. However, this hypothesis has been tested and refuted by a study by Scott and Baillargeon (2009).

In the familiarization trials of Scott and Baillargeon's (2009) false belief condition, 18-month-olds first see an agent look while a hand places a one-piece penguin and a

disassembled two-piece penguin on a platform. Secondly, infants see the agent place a key in the bottom part of the two-piece penguin before assembling the two pieces into a single piece. Thirdly, the agent leaves the scene and in her absence, a hand assembles the two-piece penguin (which is thereby indistinguishable from the one-piece penguin) and then places the assembled two-piece penguin into a transparent box and the one-piece penguin into an opaque box. In the test trials, infants see the agent reach either for the transparent or for the opaque box. Scott and Baillargeon (2009) found that infants looked longer when the agent selected the opaque rather than the transparent box, suggesting that they assumed that the agent's goal was to retrieve the two-piece penguin in order to place her key inside and ascribed to the agent the false belief that the penguin in the transparent box was the one-piece penguin and that, therefore, the two-piece penguin was in the opaque box. In the ignorance condition, everything is the same except that in the agent's absence, the hand places each penguin into an opaque box so that the agent has no (misleading) evidence about the location of the two-piece penguin. Scott and Baillargeon (2009) report that in the ignorance condition, infants looked equally when the agent selected either of the opaque boxes. This shows, contrary to the hypothesis that infants assume that ignorance leads to error, that infants did not respond in the same way to the agent's ignorance and to her false belief.

2.2. When ascribing preferences depends on ascribing false beliefs

The findings reported by Csibra, Gergely and colleagues may seem compatible with the second strategy and explainable by crediting preverbal human infants with a teleological non-mentalistic system of interpretation of an agent's action. But the findings reported by Woodward, Luo and Baillargeon raise the question whether advocates of the second strategy can coherently deny that preverbal human infants are able to ascribe false beliefs to others,

while granting them the ability to ascribe preferences to others. This question arises for both theoretical and empirical reasons.

From a theoretical standpoint, one might try to justify one's reluctance to credit infants with the ability to represent the content of an agent's false belief by arguing that, unlike an agent's true belief (or state of knowledge), an agent's false belief is *incongruent* with the state of affairs that it represents. However, while some of an agent's epistemic states (e.g. her true beliefs) are congruent with the fact that they represent, an agent's goal or intention (i.e. motivational state) is never congruent with any represented fact, because what goals and intentions represent are not facts, but *possible* (non-actual) states of affairs. In this respect, goals and intentions are like states of ignorance and false beliefs, and unlike true beliefs: they represent possible states of affairs, not actual states of affairs (or facts). If so, then advocates of the second strategy should accept that infants could represent an agent's congruent epistemic states, but find it puzzling that they could represent an agent's motivations. But so far, they have only questioned the ability of preverbal human infants to represent the contents of an agent's false beliefs, not her goals, preferences or intentions.

Furthermore, some recent evidence strongly suggests that infants ascribe a preference to an agent in accordance with the agent's epistemic states, including her false beliefs. Luo and Baillargeon (2007) compared the three following conditions: in the familiarization trials of the *transparent* condition, an agent is facing two objects (a block on the left and a cylinder on the right), both of which are visible to both the agent and the infant. In the familiarization trials of the *opaque* condition, the agent is facing the same pair of objects, both of which are visible to the infant, but only the cylinder is visible to the agent because the block is hidden from the agent's view by an opaque screen. In the *preview* condition, the familiarization trials are the same as in the opaque condition, but they are preceded by a preview trial in which the agent herself places the block behind the opaque screen, so that although the agent cannot see

the block any more, she knows of its presence behind the opaque screen. Luo and Baillargeon (2007) report that 12,5-month-olds looked longer in the test trials when the agent selected a novel target at the old position rather than the old target at the new position in both the *transparent* and the *preview* conditions, but not in the *opaque* condition. They interpret these findings as showing that infants ascribed a preference to the agent only when she was aware of the presence of two objects, which she was not in the *opaque* condition.

In the first of two further studies, Luo (2011) compared two situations, in both of which, before the test trials, an agent sits behind and in between two screens. In between the two screens there is also a block that is visible to both the agent and the infant. First, the infants see the agent place the block behind the screen on the left. Secondly, while the agent is away, the infants can see that while there is a cylinder on the right, a hand removes the block from behind the screen. But in one of the two conditions, both screens are transparent so that both the infant and the agent can see that after the hand removed the block, there is only one remaining object, i.e. the cylinder on the right. In the other condition, the screen on the left is opaque so that only the infant, not the agent, can see that after the hand removes the block, there is only one remaining object, i.e. the cylinder on the right. Thirdly, the agent is back and reaches for the cylinder on the right. Luo (2011) reports that 10-month-olds looked longer at the test trials in which the agent selects a novel object (i.e. the box) at the old location (on the right) rather than the old object (the cylinder) at a novel location (on the left) *only* in the condition in which the agent falsely believes that two objects are present (one of which is hidden from her view by the opaque screen).

In the second study, Luo (2011) tested the converse contrast in which an agent is sitting behind and in between a pair of screens. First, both the infant and the agent can see a cylinder on the right, which the agent can manipulate. Secondly, in the absence of the agent, while the cylinder is on the right, a hand places a box on the left. Thirdly, the agent comes

back and reaches for the cylinder on the right. But in one situation, only the infant, not the agent, can see the box on the left, because the screen on the left is opaque, so that the agent falsely believes that there is only the cylinder on the right. In the other situation, everything that the infant can see the agent can also see because she is sitting behind two transparent screens. In both conditions, Luo (2011) reports that 10-month-olds looked longer when the agent selected a novel object at the old location rather than the old object at a novel location, only in the situation where the agent was aware of two objects, not when she falsely believed that only one object was present. These findings strongly suggest that infants ascribe a preference to an agent and look longer when the agent fails to act in accordance with her preference, *only* if the agent correctly or incorrectly believes that she is facing two objects.[9]

In this section, I have examined the attempts by advocates of the empiricist second strategy at offering low-level alternatives to the view that preverbal human infants can represent and reason on some of an agent's false beliefs. I have found their alternative accounts unsatisfactory. Furthermore, the evidence reviewed at the end of this section strongly suggests that before they reach their first birthday, in some cases, human infants are able to ascribe motivations (e.g. preferences) to an agent only when the agent correctly or incorrectly believes that she is presented with a competing pair of objects. If so, then it is difficult for advocates of the empiricist second strategy to question the evidence suggesting that human infants can represent and reason about an agent's false belief, while taking for granted that they can represent and reason about her goals and motivations.

3. A two-systems model of belief-ascription

The first and second strategies disagree about whether or not language-understanding, pragmatic competence and executive control (necessary for inhibition), which are required to

[9] Thus, contrary to Scott and Baillargeon's (2009), Baillargeon et al.'s (2010) and Luo and Baillargeon's (2010) dichotomy between sub-system-1 and sub-system-2, before the end of their first year, infants seem able to compute an agent's false beliefs (i.e. reality-incongruent epistemic states).

pass the standard false belief task, are necessary for, or even constitutive of, the ability to ascribe false beliefs to others. Recently, Apperly and Butterfill (2009) have sketched a third intermediate strategy that seems prima facie to have the potential to reconcile the conflict between the first and the second strategies. They argue (*ibid*., p. 964) that *efficiency* and *flexibility* make competing and inconsistent demands on the ability of human adults to represent and reason about others' beliefs. To solve this tension, they argue that there are *two* systems of belief-ascription: an efficient but inflexible system (shared by human infants and adults) underlies the ascription of *belief-like* states and a flexible but inefficient system (only present in adults) underlies the ascription of *genuine* beliefs. Much of Apperly and Butterfill's (2009) argument for a two-systems model of belief-ascription rests on the assumption that no single cognitive system could be both efficient and flexible, which in turn rests on an analogy between belief-ascription and numerical cognition. Consequently, Apperly and Butterfill's two-systems model faces two related challenges: to what extent does the trade-off between efficiency and flexibility support a dissociation between two separable systems? To what extent is the two-systems model a genuine alternative to the first two strategies?

3.1. The trade-off between efficiency and flexibility

Much recent work in both comparative and developmental psychology has showed the existence of two separate so-called *core* systems for representing numerosities in both preverbal human infants and a variety of non-human animals. While the object-file system underlies the exact representation of sets of at most three individuals, the analog magnitude system underlies the approximate representation of sets of greater cardinality subject to ratio limits (of e.g. 1:2 or 2:3).[10] The expressive and operational resources of these two core systems are severely limited compared to those of the symbolic system that enables older

[10] The evidence shows e.g. that while 6-month-olds can only discriminate numerosities with a 1:2 ratio, but not with a 2:3 ratio, 10-month-old infants can do both (cf. Feigenson et al., 2004).

human children and adults to engage in mathematical reasoning. Neither core system alone can support concepts of exact integers, zero, fractions, square roots or negative numbers.[11] In Feigenson et al.'s (2004) terms, what makes numerical cognition easy are the two core systems. But as soon as one goes beyond the two core systems, numerical cognition becomes hard. On Apperly and Butterfill's (2009) interpretation, while easy numerical cognition (restricted to the two numerical core systems) seems to constitute the paradigm of cognitive efficiency, hard numerical cognition (that goes beyond the two numerical core systems) seems to constitute the paradigm of cognitive flexibility.

However, in at least three respects, the question arises whether the basic dissociation found in numerical cognition could be a good model for a two-systems model of belief-ascription. First of all, as Feigenson et al. (2004, p. 313) have insightfully argued, what drives humans beyond core numerical cognition is the existence of *two* core systems with severely limited representational resources: "if the human mind were endowed only with a single system of core knowledge, then humans might never venture beyond its bounds". Only by overcoming the limitations of each and integrating them with each other could one exactly represent the cardinality of a set whose value is beyond the representational resources of the object-file system. Only by engaging in a process of verbal counting can humans learn the exact concept expressed by a word like 'seven' whose meaning outstrips the expressive powers of either core system taken separately. But if there is a single efficient and inflexible system of belief-ascription, then it is not clear why or how humans would ever venture beyond its bounds.

Second, whereas the evidence strongly suggests that the two core numerical systems are widely shared by humans and non-human animals, there is good evidence that the system of human adults' mature arithmetic depends on literacy, not just on spoken language.

[11] See Carey (2009) and Dehaene (1997).

Furthermore, human adults in different cultures have not always used the same symbolic systems for representing natural numbers. Nor have all human adults in all cultures in the history of mankind gone beyond the limits of the two core systems (cf. Feigenson et al., 2004). In the mindreading domain, there is evidence that non-human primates and birds have the ability to represent another's visual perspective and the contrast between another's state of knowledge and ignorance. However, so far there is no convincing evidence that non-human animals can compute another's false belief (cf. Call and Tomasello, 2008). If so, then findings reporting infants' ability to represent others' false beliefs cannot support the view that it is part of a core efficient and inflexible system of belief-ascription shared by humans and non-human animals. Furthermore, there is no reason either to think that adults' flexible and inefficient system of belief-ascription depends on literacy, not just on spoken language. Nor is there evidence that some human adults in some cultures have failed to develop a full system of belief-ascription.

Thirdly, from the standpoint of human adults' mature arithmetic, the 3-item limit of the object-file system and the ratio limit of the analog magnitude system seem subject to *arbitrary* signature limitations. While the first strategy would happily grant that the efficient and inflexible system whereby preverbal human infants ascribe false beliefs to others is limited in various ways, the question arises to what extent these limitations are *arbitrary* in comparison to the resources of the inefficient but flexible system used by adults. One of the arbitrary limitations of the efficient and inflexible system considered by Apperly and Butterfill (2009, p. 957) are the infants' putative inability to fully understand the functional role of an agent's beliefs in relations to her motivations. However, both the looking time responses of infants (Onishi and Baillargeon, 2005; Surian et al., 2007; Luo, 2011) and their helping behavior (Buttelman et al., 2009) reviewed in the first two sections show that infants are sensitive to the interactions between an agent's epistemic and motivational states and,

therefore, seem hard to reconcile with the first alleged arbitrary limitation. Another arbitrary limitation is the infants' putative inability "to use all cognitively available facts to ascribe any belief that the subjects can themselves entertain" (Apperly and Butterfill, 2009, p. 964). Clearly, preverbal human infants are limited in their own belief-forming capacities: for example, they cannot acquire beliefs by means of verbal testimony. However, the evidence rather suggests that preverbal human infants can ascribe to others beliefs whose content they themselves can entertain.[12]

While Apperly and Butterfill (2009) and Butterfill and Apperly (submitted) may underestimate the flexibility of the system enabling preverbal human infants to ascribe beliefs, they may also overestimate the flexibility of the adults' system of mindreading. In fact, they seem to pretty much accept all of Davidson's (1984, 2004) picture of belief-ascription as a characterization of human adults' full system of belief-ascription.[13] Equivalently, they accept Fodor's (1983) characterization of *central* processes (i.e. *non modular*) responsible for the fixation of beliefs as being Quinean, holistic and isotropic, on the basis of which, Apperly (2011, p. 122) sketches an argument against the view that the adult system of belief-ascription could be modular in Fodor's (1983) sense. First, the argument assumes that ordinary belief-fixation is a non-modular central process (that is holistic, Quinean and isotropic). Secondly, the argument further assumes that to ascribe a belief to another is to form a belief about another's belief. The conclusion is that an ascriber's belief about another's belief could not be generated by a process that is more modular (or less central) than his or her ordinary belief-fixation mechanism. The argument is valid, but the conclusion only follows if one accepts its first premise. One can, however, reject the first

[12] One further arbitrary limitation on the efficient inflexible system hypothesized by Apperly and Butterfill (2009, p. 963) is that it does no enable infants to track beliefs whose contents involve quantifiers. I discuss this hypothesis later in section 3.3.

[13] If they do, then they presumably accept Davidson's famous claim that not unless a creature possesses the *concept* of *belief* (which, according to Davidson, requires the ability to interpret another's utterances) could she *have* beliefs. Given that Apperly and Butterfill are inclined to deny that preverbal human infants possess the concept of *belief*, the question further arises whether they are nonetheless willing to credit preverbal human infants with belief states.

premise, on the grounds that Fodor (1983) wrongly takes scientific reasoning as his model for central processes, i.e. the processes underlying ordinary belief-fixation (cf. Sperber, 1994, 1997, 2000, 2001, 2005).

3.3. How different are belief-like states from beliefs?

Whereas Apperly and Butterfill (2009) take the flexible and inefficient system to ascribe genuine beliefs, they take the efficient and inflexible system to ascribe *belief-like* states or (as the authors call them) *registrations.* The challenge is that belief-like states must share sufficiently many properties with beliefs so that they are clearly different from e.g. intentions and desires, but they also should be significantly different from beliefs for the efficient inflexible system that generates them to be distinguishable from the flexible and inefficient system that generates beliefs. The construction of the technical notion of *registration* seems primarily guided by two constraints, the first of which is that ascribing *registration* to an agent should not depart (or only minimally so) from the logical principle of *extensionality*. Secondly, ascribing *registration* to an agent should be taken not only as a condition on the agent's successful object-directed action, but also on the agent's ability to manipulate a competitor's access to a given object. It certainly is an open empirical question to what extent the cognitive life of preverbal human infants is in fact governed by the logical principle of extensionality. And so are the questions to what extent the structure of others' object-directed actions and their competitive strategies towards conspecifics are the stepping-stones for infants' efficient and inflexible mindreading capacities.

First, Apperly and Butterfill (2009) propose to define *registrations* in terms of a relation that they call the *encountering* relation, which they define on the basis of the notion of a *field* as "a region of space centered on an individual", such that "the relation obtains when the object is in the individual's field". As they make clear, the *encountering* relation is

supposed to be a purely *extensional* ternary relation holding between an agent, an object and a location (or an agent and a pair involving an object and its location). As Butterfill and Apperly (submitted, p. 12-13) observe, it is crucial to their project that the ability to track instances of the *encountering* relation does not presuppose understanding of such psychological concepts as *perceiving* and *seeing*. Indeed, for better or for worse, there are significant differences between *encountering* and *perceiving*. For example, if an agent who stands in the *encountering* relation to some object at some location has her face and eyes turned away from the object at this location, then she might hear the noise produced by the object (if any), but she will not see the object. Conversely, if her ears are blocked and the object falls in her line of sight, she might see its shape, color and texture, but she may not hear the noise produced by the object. By contrast with perception, it seems as if the *encountering* relation between an agent, a (silent) object and its location is being satisfied when the object at its location is within the agent's field, but the agent's face and eyes are turned away from the object. If so, then it is not clear that tracking the relation of *encountering* is sufficiently fine-grained to capture the sensitivity of preverbal human infants to the conditions under which an agent can or not see an object which is visible to the infants (cf. findings by Luo and Baillargeon, 2007 and Luo, 2011).

The second step is the definition of the *registration* relation based on the *encountering* relation: an agent stands in the *registration* relation to an object and its location if she *encountered* it at that location and did not encounter it elsewhere since (Apperly and Butterfill, 2009, p. 962). As I see it, the main question raised by the *registration* relation is whether a registration can be *false* or *incorrect*. I think this question generates the following dilemma for the two-systems model of belief-ascription.

A belief has a propositional content and it can be false if the state of affairs that is represented by its content fails to obtain, i.e. if it is not a fact. Now Apperly and Butterfill

(2009) and Butterfill and Apperly (submitted) strongly insist that, unlike beliefs, registrations lack propositional contents. Can a *registration* be false or incorrect? On the one hand, if one thinks of *registration* as the counterpart to *belief*, then it seems as if an agent's *registration* of an object at location p will be true (or correct) iff the object is at p at the time of evaluation and false otherwise. If so, then registrations are just like beliefs in this respect: they can be false. But if so, then registrations are unlike the *encountering* relation: they are *not* purely extensional since they represent an object at a location where it may not be. On the other hand, according to Apperly and Butterfill's (2009, p. 962) official definition, an agent S stands in the *registration* relation to object o at location p if she *encountered* o at p at some time t and did not encounter it elsewhere after t. On this construal, S's *registration* of o at p at t is not refuted by the fact that o is at a location different from p at t+1. If so, then *registrations* are like the *encountering* relation: they are extensional, but they cannot be false.

Butterfill and Apperly (submitted, pp. 24-25) further speculate that one of the *arbitrary* limitations of the efficient and inflexible system of registration-ascriptions might consist in the inability to reason about another's mistaken beliefs about an individual's *identity*. For example, on a *de dicto* (or *opaque*) reading of the proposition expressed by (1), the truths expressed by (1) and (2) do not entail that the proposition expressed by (3) is true:

(1) Mitch believes that Charly is in Baltimore.

(2) Charly is Samantha.

(3) Mitch believes that Samantha is in Baltimore.

The reason why the truth of (1) and (2) may fail to entail the truth of (3) is that Mitch may fail to know that (2) is true. Now as Butterfill and Apperly (submitted, *id*.) observe, replacement of "believe" by "register" in (1) as in (1') guarantees the truth of (3'):

(1') Mitch registers <Charly, Baltimore>

(2) Charly is Samantha.

(3') Mitch registers <Samantha, Baltimore>

If indeed preverbal human infants were ascribing registrations, not beliefs, to others, then this would explain why they fail to reason about others' beliefs about mistaken identity. But there is an alternative explanation that makes infants' inability to reason about an agent's beliefs about mistaken identity consistent with their ability to ascribe beliefs to others, when the agent's beliefs about mistaken identity involve two different names for the same individual. The explanation is simply that, so long as they cannot use English proper names, infants could not entertain either the belief that Charly is in Baltimore or that Charly is Samantha, the first of which is necessary to believe that Mitch believes that Charly is in Baltimore, and both of which are necessary to reason about Mitch's mistaken beliefs about Charly's identity.

Furthermore, there is some evidence that before they reach their second birthday, human infants can reason about others' mistaken beliefs about identity without using proper names for objects. In section 2.3.1, I discussed findings by Scott and Baillargeon (2009) showing that 18-month-olds can ascribe to an agent different representations of a two-piece penguin, one when it is disassembled and one when it is assembled and placed in a transparent box. These findings further show that 18-month-olds have different expectations about an agent's actions towards the assembled two-piece penguin, according to whether the assembled two-piece penguin is placed in a transparent and in an opaque box. These findings seem to conflict to some extent with Apperly and Butterfill's (2009, p. 957) claim that human infants cannot track beliefs about "both the features and the location of an object" or that they cannot ascend to Level 2 perspective taking (e.g. appreciating whether an agent sees an object as a two-piece penguin or a one-piece penguin).

Butterfill and Apperly (submitted, p. 29) offer an alternative interpretation of Scott and Baillargeon's (2009) findings consistent with their claims that 18-month-olds cannot reason about others' mistaken beliefs about identity. On their proposed interpretation, instead of

ascribing to the agent a mistaken belief about the identity of some *token* object, i.e. an assembled two-piece penguin in a transparent box, infants would assume that the agent is reasoning about a *type* of object. On Butterfill and Apperly's interpretation, on the basis of the familiarization trials, infants would ascribe to the agent the general expectation that she is always presented with two kinds (or types) of objects: a one piece-penguin and a disassembled two-piece penguin. Applying this general expectation to the critical test condition, the agent infers that the penguin in the transparent box belongs, not to the latter, but to the former type and further that there is (or must be) some disassembled penguin or other in the opaque box.

This is a perfectly coherent alternative interpretation of the reasoning ascribed by 18-month-olds to the agent. But this alternative interpretation raises at least two problems for advocates of the two-systems model, according to which 18-month-olds can only ascribe *registrations* to others. On the one hand, it is far from clear that an agent could ever *register* the fact that an object (or token) falls under (or belongs to) a *type* (or *kind*), in accordance with the official definition of a *registration*. Furthermore, the alternative interpretation requires that infants can ascribe to others *general*, not merely singular, beliefs, e.g. the belief that *there is some* disassembled two-piece penguin *or other* in the opaque box. If so, then this requires infants to represent and reason about beliefs whose content is an existentially quantified proposition. But this would seem to violate Apperly and Butterfill's (2009, p. 963) principle that registrations "do not permit tracking beliefs that involve quantifiers".

Thus, the advocates of the two-systems model of belief-ascription try to offer alternatives to the commonsense concepts of *perception* and *belief*, which they think reflect the signature of the arbitrary limitations displayed by findings about preverbal human infants reviewed above, while it also "involves minimizing reliance on commonsense psychological concepts in favor of a constructive account" (Butterfill and Apperly, submitted, p. 15). I have

expressed doubts about both the *arbitrary* nature of infants' limitations in belief ascription and the success of their attempt at offering alternatives to the concepts of *perception* and *belief* in terms of *encountering* and *registration*. In particular I have raised doubts as to whether *registrations* could be false (or incorrect) and extensional. In any case, like an agent's beliefs, an agent's registrations are purported epistemic states. There is much evidence that preverbal human infants keep track of both an agent's epistemic states and motivations. So advocates of the two-systems also owe us a story about the kind of motivational states which can be ascribed by the efficient and inflexible system (which they have not done so far).

Advocates of the first strategy credit preverbal human infants with the ability to ascribe jointly motivational states, whose contents are never congruent with a fact, and epistemic states, whose contents can be congruent with a fact (if it is a state of knowledge) or incongruent with a fact (if it is a state of ignorance or a false belief). They explain why it is so costly for children to succeed the standard false belief task because the latter requires not only the ability to ascribe false beliefs, but also language-understanding, pragmatic competence and executive control (required for inhibition). Now, Apperly and Butterfill (2009, p. 960) explicitly argue against the first strategy: they claim that by positing a innate capacity for representing another's false beliefs, the first strategy might "offer a solution to the problem of acquisition", but it would not thereby "explain how belief reasoning could be simultaneously cognitively efficient and cognitive flexible in adults... because the features that explain the superior abilities of adults -- greater knowledge, greater memory and executive control -- are the very features that also make adults inefficient at belief reasoning, and on this account theses features are added to the same 'one-system' that is present in infants. Innateness, then, does not, in and of itself, explain efficiency".

I simply fail to see the force of their criticism here. By positing a single system that would enable both preverbal human infants and human adults to represent others' false

beliefs, one does explain not only acquisition but also efficiency. What might make older children and adults not only more flexible, but also less efficient, than preverbal human infants is precisely reliance on greater memory, language-understanding, pragmatic competence and executive control: combining belief-ascription with these further resources might increase flexibility at the cost of efficiency. If so, then success at elicited false belief tasks that require the ability to represent an agent's false belief to be supplemented by greater memory, language-understanding, pragmatic competence and executive control would be evidence of greater flexibility acquired at the cost of efficiency and perhaps also automaticity. If so, then the trade-off between efficiency and flexibility might not be sufficient for grounding the duality between two systems of belief-ascription.

Concluding remarks

This paper turns out to be an argument for the first strategy: this strategy takes findings about the ability of preverbal human infants based on the violation-of-expectation paradigm at face value as evidence that they can ascribe false beliefs to others. The first strategy further accounts for why it is so hard for children until the end of their fourth year to pass the elicited false belief task because the latter requires several further cognitive capacities in addition to the ability to ascribe false belief. The second strategy assumes that the ability to pass the standard false belief task is a criterion for the ability to ascribe false beliefs to others. It is therefore incumbent upon advocates of the second strategy to offer deflationary accounts of the findings about preverbal human infants without crediting them with the capacity to ascribe false beliefs to others. In the second section, I have argued that three of the most important deflationary accounts proposed by advocates of the second strategy are inadequate. In the last section of the paper, I have examined a recent version of a

two-systems model of belief-ascription and I have argued that it fails to constitute a genuine alternative to the second strategy, which I have found unsatisfactory.

References

Apperly, I. (2011) *Mindreaders, the Cognitive Basis of "Theory of Mind"*, New York: Psychology Press.

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

Baillargeon, R., Scott, R.M. and He, Z. (2010) False-belief understanding in infants. *Trends in Cognitive Sciences*, 14, 3, 110-118.

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

Bennett, J. (1978) Some remarks about concepts. *Behavioral and Brain Sciences*, 4, 557-560.

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

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

Butterfill, S. and Apperly, I. (submitted) How to Construct a Minimal Theory of Mind.

Call, J. and Tomasello, M. (2008) Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12, 5, 187-192.

Carey, S. (2009) *The Origin of Concepts*, Oxford: Oxford University Press.

Caron, A.J. (2009) Comprehension of the representational mind in infancy. *Developmental Review*, 29, 69–95.

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

Csibra, G., Gergely, G., Bíró, S., Koós, O. and Brockbank, M. (1999) Goal attribution without agency cues: The perception of ‘‘pure reason” in infancy. *Cognition,* 72, 237–267.

Davidson, D. (1984) *Truth and Interpretation*, Oxford: Oxford University Press.

Davidson, D. (2004) *Problems of Rationality*, Oxford: Oxford University Press.

Dehaene, S. (1997) *The number sense: How the mind creates mathematics*, Oxford: Oxford University Press.

Dennett, D.C. (1978) Beliefs about beliefs. *Behavioral and Brain Sciences,* 4, 568-570.

Feigenson, L., Dehaene, S. and Spelke, E.S. (2004) Core systems of number. *Trends in Cognitive Sciences*, 8, 7, 307-314.

Fodor, J.A. (1983) *The Modularity of Mind*, Cambridge, MA: MIT Press.

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

Gergely, G. and Csibra, G. (2003) Teleological reasoning about actions: the naïve theory of rational actions. *Trends in Cognitive Sciences*, 7, 287-292.

Goldman, A. (2006) *Simulating Minds, the philosophy, psychology and neuroscience of mindreading*, Oxford: Oxford University Press.

Harman, G. (1978) Studying the chimpanzee’s theory of mind. *Behavioral and Brain Sciences,* 4, 576-577.

Hogrefe, G.-J., Wimmer, H. and Perner, J. (1996) Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development,* 1986, 57, 567-582.

Jacob, P. (in press) Sharing and ascribing goals. *Mind & Language*.

Jacob, P. and Jeannerod, M. (2003) *Ways of Seeing, the Scope and Limits of Visual Cognition*, Oxford: Oxford University Press.

Leslie, A.M. (2000) ‘Theory of mind’ as a mechanism of selective attention. In M. Gazzaniga (ed.) *The New Cognitive Neurosciences*, 2nd Edition (pp. 1235–1247), Cambridge, MA: MIT Press.

Leslie, A.M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, 9, 459–462.

Luo, Y. and Baillargeon, R. (2005) Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, 16, 601-608.

Luo, Y. and Baillargeon, R. (2007) Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, 105, 489-512.

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

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

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

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

Premack, D. and Woodruff, G. (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4, 515-526.

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. and Perner, J. (2005) Do infants really understand false belief? Response to Leslie. *Trends in Cognitive Sciences*, 9, 462– 463.

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

Scott, R.M., Baillargeon, R., Song, H.-J. and Leslie, A. (2010) Attributing false beliefs about non-obvious properties at 18-months. *Cognitive Psychology*, 61, 4, 366-395.

Searle, J.R. (1983) *Intentionality, an Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press.

Southgate, V., Senju, A. and Csibra, G. (2007) Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18, 7, 587-192.

Sperber, D. (1994) The modularity of thought and the epidemiology of representations. In L. Hirschfeld and S. A. Gelman (eds.) *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 39-67), Cambridge: Cambridge University Press.

Sperber, D. (1997) Intuitive and reflective beliefs. *Mind and Language, 12*(1), 67-83.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (ed.) *Metarepresentations: A Multidisciplinary Perspective* (pp. 117-137), Oxford: Oxford University Press.

Sperber, D. (2001) In defense of massive modularity. In E. Dupoux (ed.) *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler* (pp. 47-57), Cambridge,

MA: MIT Press.

Sperber, D. (2005) Modularity and relevance: How can a massively modular mind be flexible and context-sensitive? In P. Carruthers, S. Laurence and S. Stich (eds.) *The Innate Mind: Structure and Contents*, Oxford: Oxford University Press.

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

Warneken, F. and Tomasello, M. (2006) Altruistic helping in human infants and young chimpanzees. *Science*, 311, 1301–1303.

Warneken, F. and Tomasello, M. (2007) Helping and cooperation at 14 months of age. *Infancy*, 11(3), 271–294.

Wimmer, H. and Perner, J. (1983) Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128.

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

Woodward, A.L. (1999) Infants' ability to distinguish between purposeful and nonpurposeful behaviors. *Infant Behavior and Development*, 22, 145–160.