The developmental paradox of false belief understanding: a dual-system approach

Leon de Bruin[1] and Albert Newen[1]

[1] Ruhr-University of Bochum, Bochum, Germany
lcdebruin@gmail.com
albert.newen@rub.de

Abstract. In our article we explore the developmental paradox of false belief understanding. This paradox follows from the claim that young infants already have an 'implicit' notion of false belief, despite the fact that they consistently fail tests assessing 'explicit' forms of false belief understanding. First, we argue that recent dual-system proposals to solve this paradox are unsatisfactory because they either lack the conceptual resources to deal with the differences between implicit and explicit false belief understanding, or ignore questions about system interaction. Second, we discuss a number of problems dual-system approaches have to address in order to account for the development of false belief understanding, and propose a model that combines a layered model of perspective taking with an inhibition-selection-representation mechanism operating on different levels.

Keywords: Theory of Mind, False Belief Task, implicit false belief understanding, shared representations, decoupling

1. Introduction

In our everyday social interactions we frequently attribute mental states (e.g., beliefs, desires) to others in order to predict or explain their behavior. For example, if Mini knows that Maxi wants the chocolate, and she also knows that Maxi believes the chocolate is in the blue cupboard, she can predict that Maxi will search for the chocolate in the blue cupboard. This works equally well if Mini wants to explain Maxi's behavior: Maxi's searches in the blue cupboard, because he wants the chocolate and believes the chocolate is in the blue cupboard. It is generally accepted that this capacity is enabled by a Theory of Mind (e.g., Perner 1991, Baron-Cohen 1995, Leslie 2000).

Over the last couple of decades, most research in Theory of Mind has focused on the development of false belief understanding. In order to test whether Mini truly understand Maxi's behavior from *his* point of view, we need to make sure that she does not simply attribute to Maxi her *own* belief about the location of the chocolate. The False Belief Test (FBT) has been specifically designed to solve this problem by introducing a condition in which the protagonist has a *false* belief about some state of affairs in the world (usually the location of an object).

In 'explicit' (i.e., verbal) versions of this test, children are asked a direct question about the protagonist's false belief. In the 'unexpected location' FBT (Wimmer & Perner 1983, Baron-Cohen et al. 1985), for example, children observe a protagonist who sees an object being placed in a certain location. Then the protagonist leaves, and the object is moved to another location. When the protagonist returns, he mistakenly believes the object is still in its initial location. At this point, the children are asked to predict where the protagonist will look for the object.

Test results show that 3-year-olds typically give a wrong answer to this question, while four-year-olds answer correctly. Findings on other explicit FBTs, such as the 'unexpected identity' task (Perner et al. 1987; Moses & Flavell 1990; Wellman 1990), confirm this picture. Many researchers have therefore concluded that false belief understanding does not emerge until four years (Flavell 2004, Sodian 2005; see Wellman 2002 for a review, and Wellman et al. 2001 for a meta-analysis).

Recently, however, this conclusion has been challenged on the basis of findings on so-called 'implicit' false belief understanding. Studies based on 'violation of expectation' and 'anticipatory looking' paradigms have claimed that implicit false belief understanding is already present at a considerably earlier age, in 25-month-olds (Southgate et al. 2007), 15-month-olds (Onishi & Baillargeon 2005), and even 13-month-olds (Surian et al. 2007). These findings give rise to a very interesting and widely debated 'developmental paradox': if young infants already understand false belief, as the above studies seem to suggest, then why do they fail the explicit FBT?

Our aim in this article is to propose a model that allows us to account for both implicit and explicit forms of false belief understanding and to shed new light on the developmental paradox. We start by discussing a number of important studies on implicit false belief understanding (section two). Next, we evaluate two recent dual-system accounts of the developmental paradox (Apperly & Butterfill 2009; Baillargeon et al. 2010) in section three. We argue that both accounts fail to solve the developmental paradox because they either lack the conceptual resources to deal with the differences between implicit and explicit forms of false belief understanding, or ignore questions about the interaction between the two systems. In section four, we discuss a number of problems dual-system approaches have to address in order to account for the development of false belief understanding. Section five puts forward a model combining a layered model of perspective taking with an inhibition-selection-representation mechanism that operates on different levels.

2. Implicit False Belief Understanding

The explicit FBT places rather strong demands on children's other cognitive capacities (Bloom & German 2000, Carlson & Moses 2001). Studies on implicit false belief understanding are designed in a way as to reduce these demands in order to see whether children might be capable of false belief understanding at an earlier age. In the implicit FBT, infants no longer have to give an explicit answer to a question about the protagonist's belief. Instead, their understanding of false belief is inferred from the behavior they spontaneously produce (cf. Baillargeon et al. 2010).

Clements & Perner (1994) already showed that linguistic competence is responsible for at least some of the difficulties infants have with the explicit FBT vis-à-vis its implicit counterpart. They adapted an early version of the 'anticipatory looking' paradigm, which is used to test whether children are able to visually anticipate where another agent will search for an object, given his false belief about its location. In their experiment, Clements & Perner (1994) asked children to watch how the protagonist, a mouse called Sam, stored an object in a box in front of one of two mouse holes. While Sam is asleep, the object is moved to a different box in front of the other mouse hole. The experimenters monitored where the children were looking in anticipation of Sam's reappearance, given his false belief about the object's location. They contrasted this spontaneous behavior with the explicit answers children gave to the question where Sam would look for the object.

Clements & Perner (1994) found that 3-year-olds looked at the initial (correct) location when anticipating Sam's return, even when they explicitly made the incorrect claim that he would go to the second location. They labeled this early manifestation of false belief understanding 'implicit', because the participating children were not explicitly aware of the knowledge conveyed in their correct eye gaze. Importantly, the researchers found no sign of implicit false belief understanding in 2-year-old children (cf. Perner & Clements 2000, Clements & Perner 2001, Ruffman et al. 2001, Garnham & Perner 2001).

However, recent studies on implicit false belief understanding challenge the assumption that infants under three do not understand false belief. Onishi & Baillargeon (2005), for example, used a 'violation-of-expectation' FBT to investigate whether children would look reliably longer when agents act in a manner that is inconsistent with their false beliefs. In the experiment, 15-month-old infants were familiarized with a protagonist hiding a toy in one of two locations. The protagonist left, and the toy was moved without his knowledge. Then the infants were shown scenes of the protagonist searching for the hidden toy, either where he falsely believes it to be, or where it was actually located. Onishi & Baillargeon (2005) found that infants looked significantly longer at those scenes in which the protagonist searched at the correct location despite their false belief about where the toy was hidden. Follow-up experiments have shown similar results in even younger infants of thirteen months (Surian et al. 2007).

These findings contradict the results of Clements & Perner (1994). According to Southgate et al. (2007), these different results are due to the fact that Clements & Perner (1994)'s experiment still included a verbal element: in order to maximize the frequency of anticipatory looking at one of the mouse holes, the investigator said aloud, 'I wonder where Sam is going to look?' before asking the question. Southgate et al. (2007) argue that this primed 2-year-old infants to look at the incorrect location. In their own study, they removed the verbal element from the design and used an eye-tracker to measure anticipatory looking in 25-month-olds. The infants observed how a protagonist witnesses a puppet bear that hides a ball in one of two boxes. Then the protagonist becomes distracted and turns away from the scene. Meanwhile, the bear removes the ball from its original hiding location. Southgate et al. (2007) found that most 25-month-olds correctly anticipated the protagonist's behavior and looked at the location where she falsely believed the ball to be hidden.

Other implicit FBTs indicate that infants do not only understand false beliefs about locations, but also about number, identity, and other properties (He & Baillargeon 2007, Scott & Baillargeon 2009, Scott et al., 2007, Song & Baillargeon 2008). These findings suggest that, contrary to what the results of explicit FBTs suggest, false belief understanding may already be present by the age of 13-18 months (see Baillargeon et al. 2010, Poulin-Dubois et al. 2009 for reviews) and perhaps even earlier (e.g., Kovács et al. 2010).

3. The Developmental Paradox of False Belief Understanding: Two Recent Solutions

Whether or not these findings should be interpreted under the heading of *false belief* understanding has been a weighty topic of discussion (Perner & Ruffman 2005; Ruffman & Perner 2005; Csibra & Southgate 2006; Sirois & Jackson 2007; Herschbach 2008). The issue is important because an early onset of false belief understanding during the first year suggests that the core principles of our Theory of Mind are largely part of our biological inheritance, whereas an onset at four years makes it more plausible that Theory of Mind is influenced by cultural processes and closely tied to language acquisition.

If false belief understanding already emerges during the second year of life, then the crucial question is why 3-year-old children consistently fail the explicit false belief task (Wellman et al. 2001), even when paradigms are used that reduce response selection and inhibition demands (e.g., Call & Tomasello 1999; Sodian et al. 2006). This is what we call the 'developmental paradox' of false belief understanding.

Most participants in the debate agree that there is no simple solution to the developmental paradox. Some of them have argued that we need two different systems to account for the intricacies of false belief development (e.g., Call & Tomasello 2008, Penn et al. 2008). Apperly & Butterfill (2009), for example, propose that the findings on implicit and explicit false belief understanding are best explained by respectively an early 'Minimal' Theory of Mind, which is cognitively efficient but limited and inflexible, and a flexible but cognitively very demanding full-blown Theory of Mind. Moreover, they offer an interpretation of infants' performance on the implicit FBT in terms of 'belief-like' states rather than full-blown beliefs. On Apperly & Butterfill's interpretation, infants are sensitive to the agent's belief only insofar as they *register* the object. First, however, Apperly & Butterfill explain the simpler notion of *encountering*. Encountering is defined as 'a relation between an individual, an object and a location, such that the relation obtains when the object is in the individual's field' (p. 962). A field is defined, simply, as a certain region of space around the individual. Building on this, registering is defined as a slightly more complex psychological relation that obtains between an individual, an object and a location. An individual is said to register an object at a location when (a) she encounters the object at the location and (b) has not since encountered it somewhere else. A registering is off target when the object registered is not located where it is registered to be. The importance of the concept lies in the connections to actions: 'One can understand registration as an enabling condition for action, so that registering an object and location enables one to act on it later […] Further, registration also can be understood as determining which location an individual will direct their actions to when attempting to act on that object' (962). Tracking an agent's registration of something does not require sensitivity to her mental states as propositional attitudes – which only comes with a full-blown Theory of Mind.

Apperly & Butterfill's (2009) proposal seems to provide us with a promising way to explain the different forms of false belief understanding. However, it also raises an important question: how are the two Theory of Mind systems (functionally) related? One possibility is that there is no direct interaction between the early-developing and later-developing Theory of Mind systems. Apperly & Butterfill (2009) think this assumption is supported by the afore-mentioned findings of Clements & Perner (1994) and Southgate et al. (2007). For they show that infants correctly anticipate the action of another agent in terms of looking behavior, even when they respond incorrectly when asked to make an explicit prediction about the agent's future action. Apperly & Butterfill (2009) claim that these results are consistent with the possibility that an early-developing Theory of Mind system for tracking belief-like states is guiding children's eye movements and a later-developing Theory of Mind system underlies children's explicit judgments about beliefs.

The claim that there is no interaction between the two systems *at all* strikes us as implausible, however. For this would imply that our full-blown Theory of Mind system remains invisible during the first years of development and suddenly becomes fully operational at age four. At least we would expect to see *precursors* to such a Theory of Mind system that contribute to the system's functioning. For example, although Baron-Cohen's (1995) Theory of Mind Mechanism can be understood as a separate, late-developing mechanism, it depends on and receives input from the earlier developing Intentionality Detector, Eye-Direction Detector, and Shared Attention Mechanism.

Moreover, recent studies (e.g., Aschersleben et al. 2008, Sodian et al. forthcoming) indicate that the set of specific competencies required for the implicit FBT (e.g., looking-time patterns, eye-direction detecting, etc.) actually predicts performance on the explicit FBT. While this does not imply that there has to be interaction 'all the way down' or 'all the way up' between the two ToM systems, it seems to us that a complete theoretical account has to explain not only the dissociations but also the continuations in the development of false belief understanding. Therefore, we shall assume that the later-developing, cognitively more demanding ToM system depends, at least to some extent, on the early minimal ToM system for its operation (cf. Csibra & Gergely, 1998; Russell, 2007; Surian et al., 2007).

Baillargeon et al. (2010) have recently proposed a different solution to the developmental paradox. According to their proposal, infants come equipped with a psychological reasoning system that consists of two subsystems: sub-system 1 and sub-system 2. Sub-system 1 enables infants to attribute both motivational states and reality-congruent informational states to other agents, and is well in place by the end of the first year. Motivational states are defined as states that specify the agent's motivation in the scene and include goals and dispositions. Reality-congruent informational states, by contrast, specify what knowledge or accurate information the agent possesses about the scene. Sub-system 2 deals with reality-incongruent informational states, i.e. false beliefs, and becomes operational in the second year of life.[1]

Baillargeon et al. (2010) argue that the developmental paradox of false belief understanding can be solved by means of a careful analysis of the task-requirements of explicit and implicit FBTs. They claim that, whereas the implicit FBT only involves (i) a process of false belief representation, the explicit FBT also requires (ii) a response selection process (when asked the test question, children must access their representation of the agent's false belief to select a response) and (iii) a response-inhibition process (when selecting a response, children must inhibit any prepotent tendency to answer the test question based on their own knowledge (see also Scott and Baillargeon 2009).

What is problematic about Baillargeon et al. (2010)'s solution is that it seems arbitrary to argue that the implicit FBT only involves false belief representation, whereas its explicit counterpart also require selection processing and response-inhibition. If we accept that the implicit FBT involves false belief representation, then it is not clear why it does not require selection processing and response-inhibition as well. For infants still have to select a false belief among other beliefs, and in order to do so they have to inhibit their default attribution of true beliefs.[2] And if this is correct, then it does not work to argue that the failure on the explicit FBT is due to the joint activation of false-belief-representation processes and response-selection processes, which 'overwhelms' the child's limited information-processing resources, and/or the fact that the neural connections between the brain regions that serve these two processes are still immature and inefficient in early childhood.

In other words, whereas Apperly & Butterfill (2009) ignore questions about the interaction between the two systems involved in false belief understanding, Baillargeon et al. (2010) seem to lack the conceptual resources to adequately explain the differences between implicit and explicit forms of false belief understanding.

4. Some Requirements for a Dual-system Account of False Belief Understanding

Most accounts of explicit false belief understanding assume that infants have a default tendency to attribute their own (true) beliefs to other agents (e.g. Leslie et al. 2004) or to respond on the basis of their own knowledge (Birch & Bloom 2007; Carlson & Moses 2001). In order to pass the explicit FBT, infants have to be capable of taking 'offline' (i.e. inhibiting) their own reality-congruent perspective.

According to Baillargeon et al. (2010), as we saw in the previous section, the ability for offline processing is precisely what constitutes the difference in task demands between the implicit and explicit FBT. However, this seems to ignore the fact that the implicit FBT involves offline processing as well, albeit it of a less demanding kind. Take the 'violation-of-expectation' study by Onishi & Baillargeon (2005), for example (see section 2). Although this experiment does not require infants to deal *explicitly* with differences in belief, it does require them to process differences between the visual information available to themselves and the visual information available to the other agent. This can only be accomplished offline, since the other's visual information is not directly available to the infant and needs to be represented by her. Therefore, already the implicit FBT involves a capacity for decoupling from one's own online processing of visual information and processing offline a representation of the visual information accessible to another agent. Yet, the role of decoupling and offline processing in this study is still limited. The infant largely relies on online visual information, and only has to process offline the other

---

[1] This proposal is an extension of Leslie's (1994) view that the ToMM (Theory of Mind Mechanism) consists of two subsystems: sub-system 1, which is available from 6 to 8 months of age and allows for the processing goal-directed actions, and sub-system 2, which is available from approximately 18 months of age and generates propositional attitude representations.

[2] Baillargeon et al. (2010) assume that infants by default attribute true beliefs to other agents.

agent’s representation of the location of a *single* object. More difficult versions of the implicit FBT place stronger demands on offline processing. For instance, Song & Baillargeon (2008) conducted an experiment in which infants had to represent the visual representation of another agent with respect to *both* the location and the identity of *two* objects. Among implicit FBTs, we can thus distinguish between more or less demanding versions requiring more or less decoupling.

The above analysis shows that postulating a single sub-system, such as Baillargeon et al. (2010)’s sub-system 2, fails to explain the differences between implicit and explicit forms of false belief understanding.[3] This is because both implicit and explicit false belief understanding seem to require a three-step inhibition-selection-representation mechanism. Thus, an additional system is needed to explain why young infants already have an implicit understanding of false belief, but consistently fail tests assessing explicit false belief understanding. At the same time, contrary to what Apperly & Butterfill (2009) claim, this alternative dual-system account also has to say something about how the two systems interact throughout the development of false belief understanding.

In order to determine what is required of such a dual-system account it pays to consider an early dual-system view proposed by Alan Leslie. According to this view, explicit false belief understanding depends on (i) a Theory of Mind Mechanism, and (ii) a Selection Processing System. The Theory of Mind Mechanism contains the basic meta-representational concept of belief (as well as the concept of desire and pretense) and provides the infants with an early intentional insight into the behavior of others. Leslie assumes that infants have a default strategy of attributing their own true belief to others.[4] In order to pass the explicit FBT, they must learn to inhibit or override this default strategy, and select the content of the other agent’s false belief. Leslie argues that both processes are handled by the Selection Processing System. The job of this system is essentially to ‘inhibit competing possible contents for the belief’ (Scholl & Leslie 1999, p. 147). Moreover, ‘to infer the content of somebody's belief when that content is false, SP is required to select among the possible contents that ToMM makes available’ (ibid.). However, since the Selection Processing system matures later than the Theory of Mind Mechanism (which is innate), the infant fails the false belief task until the Selection Processing system is in place. ‘The developing performance on false-belief tasks, on this view, reflects not a developing ToMM, but a developing SP’ (ibid., p.149).

Is such a ToMM/SP dual-system view able to account for implicit forms of false belief understanding? Interestingly, Leslie (2005) does mention the findings by Onishi & Baillargeon (2005). He claims that they ‘underline the early role of ToMM as a core mechanism of attention, identifying learning opportunities as expectations are violated and directing attention to relevant sources of information’ (p.532). According to Leslie, the experiment by Onishi & Baillargeon (2005) and others reflect a very early understanding of false belief, which means that SP is already operational in infants of 15 months. However, if this is true, then the whole idea that infants use a default attribution strategy before they pass the explicit FBT becomes problematic. For it means that SP initially provides infants with an early implicit understanding of false belief, then, at a later point of development, exchanges this for a default attribution strategy (despite the fact that the infant already has an understanding of the difference between true and false belief), only to reinstall an explicit form of false belief understanding around the age of four. Nevertheless, Leslie (2005) does seem to follow this line of thought, for he speculates that ‘sometime between 15 and 30 months, SP learns to make the true-belief attribution the default’ (p.532).

---

[3] Although Baillargeon et al. (2010) postulate two sub-systems, only *one* of them (sub-system 2) is used to explain the difference between implicit and explicit false belief understanding.

[4] Leslie (2000) explains the importance of default belief attribution as follows: 'it is useful to understand why belief attribution has a default bias. If desires set an agent's goals, beliefs inform the agent about the state of the world. A belief that misinforms an agent is a useless, even a dangerous thing: beliefs ought to be true. Therefore, the optimal default strategy for the belief attributer is to assume that an agent's beliefs are true.’ (p.1242)

This is puzzling. Why would SP make true-belief attribution the default when infants already have an understanding of false belief, especially given the fact that it has to be overridden again at a later stage of development? The problem here seems to be that Leslie tries to explain the development of false belief understanding completely in terms of the development of a single Selection Processing System, much like Baillargeon et al. (2010) try to explain it completely in terms of the inhibition-selection-representation processes enabled by sub-system 2.

We think that what is needed instead is a more dynamic view of the *concepts* our Theory of Mind Mechanism makes available and how they unfold throughout ontogeny (in close interaction with an inhibition-selection-representation mechanism). This is not per se incompatible with the idea of innate mental state concepts. Spaulding (2010), for example, maintains that innate mental state concepts are not necessarily 'robust' (p.127). She contrasts robust mental states, typified by propositional attitudes such as belief, with 'sub-doxastic' mental states (p.123) that do not possess truth-evaluable, propositional content.[5]

This suggestion is very much in line with recent proposals that emphasize the need to get away from the standard folk psychological concepts of belief and desire when it comes to explaining early socio-cognitive capacities such as those recruited for implicit false belief understanding. These proposals range from the postulation of a Naïve, Weak, or Minimal Theory of Mind, i.e., one lacking paradigm folk psychological concepts (Bogdan 2009, Tomasello et al. 2003, Apperly & Butterfill 2009), Perceptual Mindreading (Bermúdez 2009), or an Early Mindreading System (Nichols & Stich 2003).

We already mentioned Apperly & Butterfill's (2009) proposal of an early-developing Minimal Theory of Mind, which enables infants to track belief-like states and guides their eye movements. In a similar vein Bogdan (2009) postulates the existence of a Naïve Theory of Mind which has 'the primary function of registering and representing another mind's *relations* to the world' (p.63). Naïve Theory of Mind is best understood as an assembled cluster of abilities that enables the grasping and representing of the mental states of others, specifically - gazing, seeing, and emoting. Bogdan (2009) proposes that infants slowly move from initially noticing such things as another's direction of gaze, bodily posture or movement in purely *behavioral* ways to being able to track, register, or represent the target of the *purposed aboutness* of another's attending (2009, p.71).

Like Apperly & Butterfill's (2009) 'encountering' and 'registering', however, it is not clear whether this notion is sufficiently fine-grained to properly explain the different abilities involved in various forms of false belief understanding. Bogdan (2009) explains 'purposed aboutness' as a kind of goal-directed intentionality that is recognizably expressed in the way that organisms respond to aspects of their immediate environment - showing (1) *relatedness* to a target; (2) the *direction* of this relatedness; and (3) the *target* itself. But such an explanation is too general and fails to capture important differences between the abilities required for false belief understanding. Consider Southgate et al. (2007), for example, who employed an eye-tracker to measure anticipatory looking in 25-month-olds. In the experiment, infants observed how a protagonist witnesses a puppet bear that hides a ball in one of two boxes. Then the protagonist becomes distracted and turns away from the scene. Meanwhile, the bear removes the ball from its original hiding location. Southgate et al. (2007) found that most 25-month-olds correctly anticipated the protagonist's behavior and looked at the location where she falsely believed the ball to be hidden. Like the 'violation-of-expectation' study by Onishi & Baillargeon (2005), this experiment requires infants to process differences between the visual information available to themselves and the visual information available to the protagonist. Since the latter is not directly available, it has to be represented by the infant. In this scenario, an understanding of the agent's *purposed aboutness* comes down to being able to anticipate her behavior on the basis of her *perception* of a given object in the previous scene.

Now take a different experiment by Luo & Baillargeon (2005). This study showed that infants of 5 months old, after watching an agent repeatedly reach for object A as opposed to object B, registered the

[5] See Stich (1978) for the origins of this distinction.

‘purposed aboutness’ (i.e. preference) of this agent for object A over object B. When the objects’ positions were reversed, infants expected the agent to reach for object A in its new position, and they looked reliably longer if the agent reached for object B instead. What is important is that in this experiment, unlike the one by Southgate et al. (2007), infants did not need to process *perceptual* incongruencies between their visual perspective and that of the other agent. Both infant and agent perceived the same object, and the infant does not need to represent the reality-incongruent informational state of another agent. In this experiment, understanding *purposed aboutness* means being able to anticipate the behavior of another agent on the basis of her *movements* towards the object.

Both behavior-anticipation strategies depend on different abilities, and although both emerge in the first year of life, they have a different developmental onset. Visual habituation studies indicate that infants are already capable of understanding another agent’s goal-directed movement towards an object from 5 months onwards (e.g., Biro & Leslie, 2007; Gergely & Csibra, 2003; Woodward, 1998, 2005). Importantly, they do not selectively attend to goals for events involving inanimate objects, such as rods or claws (Woodward, 1998), or for events in which the agent’s hand is disguised by a metallic glove (Guajardo & Woodward, 2004).[6] Understanding another’s agent goal-directed perception emerges later in development. In an experiment by Woodward (2003), for example, 7- and 9-month-old infants followed an agent’s gaze and when they saw the agent look at and grasp a toy, they not only looked at that toy, but also selectively registered the directedness of the agent towards the toy. However, when the infants only saw the agent look at the object but not touch it, they failed to register this directedness. Infants of 12 months, in contrast, were capable of registering the agent’s directedness towards the toy solely on the basis of her gaze. Other studies also indicate that it is only towards the end of the first year that infants become capable of registering more abstract (or ‘distal’) goal-directed behaviors such as looking and pointing (Phillips et al. 2002, Sodian & Thoermer 2004, Woodward 2003, 2005, Woodward & Guajardo 2002).

This shows that the concept of purposed aboutness is not fine-grained enough to capture the different abilities recruited in the various implicit FBTs. The same seems to hold for the notions of ‘encountering’ and ‘registering’ proposed by Apperly & Butterfill (2009). Of course, this is not a principled objection against these positions. Rather, it should be seen as an incentive to develop a more sophisticated conceptual vocabulary to do justice to the empirical findings on false belief understanding.

Ideally, this conceptual vocabulary should be sufficiently flexible to account for two important developmental interactions, namely, (i) between the infant’s own action towards an object and her subsequent perception of another agent’s goal-directed behavior towards the object, and (ii) between the infant’s perception of another agent’s goal-directed behavior towards an object and her own perception of the object.

With respect to (i): a study by Sommerville et al. (2005) demonstrated that even 3-month-olds focus on the relation between an actor and her goal if they reached (and not just watched) for a toy before observing another agent grasping it. The more they themselves were engaged in object-directed contact with the toys, the more sensitive they were to the agent goal-directed behavior. More recently, Sommerville et al. (2008) also found that 10-month-old infants who received active training in pulling a cane to retrieve a toy subsequently registered another person’s cane-pulling actions as goal-directed behavior, while infants who underwent observational training were unable to do this.

---

[6] Early visual habituation experiments showed that infants did not register the goals of inanimate objects (Woodward 1998, Guajardo & Woodward, 2004). Recent findings, however, suggest that infants do sometimes perceive inanimate entities as goal-directed agents (Biro & Leslie, 2007; Csibra, 2008; Johnson et al. 2001; Kuhlmeier et al. 2003; Mahajan & Woodward 2009; Luo & Baillargeon, 2005; Shimizu & Johnson, 2004). This seems to depend on the availability of additional cues (e.g. self-propelled motion) indicating the animacy of the agent. The current debate is mainly concerned with the range of cues that might contribute to the infants’ goal understanding (cf. Biro & Leslie, 2007).

With respect to (ii): it has been shown that behavioral cues (Biro et al. 2007, Biro & Leslie 2007), infant-directed talk and eye-contact induce gaze-following towards specific objects (Senju & Csibra 2008), and the use of specific linguistic labels increases the salience of objects over others (Xu 2002, Xu et al. 2004). Other experiments suggest that reaching behavior promotes the infant's perception of the spatiotemporal properties of the object of interest, whereas pointing behavior promotes the surface properties of the object (e.g., Csibra & Gergely 2006).

This raises important questions about how the infant's own action informs her perception of goal-directed behavior and vice versa. Some have proposed that the actions of infants and their perceptions of the actions of others are by a cross-modal system that translates action and perception into a unified 'language'. Georgieff & Jeannerod (1998) proposed the term 'shared representation' in order to articulate the idea that action and perception might essentially share the same representational space. This possibility is compatible with the findings of mirror neurons – neurons that fire during both action production and action perception (e.g., Rizzolatti & Craighero 2004, Rizzolatti et al. 2006). It has been argued that these mirrors show a 'human bias', in the sense that they resonate stronger with perceived actions of human versus non-human agents (Press et al. 2007, Tsai et al. 2008). This link between perception and action has also been found in early infancy (Kanakogi & Itakura 2010), and research on newborn imitation has been cited as evidence for an inherited mirror neuron system that underlies imitative behavior in human infants (e.g., Iacoboni et al.1999; Decety et al. 2002; Grezes et al. 2003; Iacoboni 2005; Iacoboni & Depretto 2006). However, there are many open questions about the existence of such a system in infants and its role in infant development (Gerson & Woodward in press; Meltzoff 2006). Moreover, there are also methodological doubts (Hickok 2009) and experiments that fail to report mirror neuron activity (Lingnau et al. 2009).

Therefore, it should be emphasized that mirror neurons are just one way of getting at the more general idea that action production and action perception can be understood in terms of shared representations. There are other grounds for supporting this idea as well, e.g. on the basis of proposals about action coding (e.g., Elsner & Hommel 2001, Prinz 2002) or findings from developmental studies (Meltzoff 2004, 2006; Meltzoff & Moore 1977, 1994; Meltzoff & Brooks 2001).

5. False Belief Understanding as Progressive Decoupling

What is attractive about shared representations is that they explain how perception and action are dynamically co-constituted in what Gallese (2001) calls the primordial 'we space'. However, this also gives rise to an important question about the registration of *agency*. How are the mechanisms that facilitate shared representations able to differentiate situations in which infants observe the goal-directed behavior of another agent from those in which they perform the same action themselves – such as those in the studies by Sommerville et al. (2005, 2008)? This is a serious problem for those who appeal to mirror neurons in their explanation of action understanding. Since both conditions activate the same cortical 'mirror' sectors, an additional mechanism is needed to determine whether the infants performs or observes the action. More in general, it can be seen a problem for proponents of a version of Simulation Theory. The question is, as Gordon (1986) puts it, how infants manage to make 'adjustments for the relevant differences' while avoiding 'total projection'.

Although we think this is indeed an important problem, it should not be overstated. To start with, researchers have proposed various solutions to address this issue. According to one proposal, shared representations are neither first- or third-person. The infant's observation of goal-direct behavior triggers the activation of so-called 'naked representations'. The idea is that mirror neurons encode the sensorimotor and perceptual properties of goal-directed behavior (either perceived or produced) in a shared representational format, but do not register the agent behind the action (deVignemont 2004; Jeannerod & Pacherie 2004; Gallese 2005; Hurley 2008). This is done in a second step by an additional mechanism. Georgieff & Jeannerod (1998), for example, have argued that this process might be taken care of by a 'Who' mechanism. Evidence for this mechanism comes from experiments showing a differential activation in the posterior insula when the subject took the role of agent, and in the right inferior parietal cortex when it took the role of observer (Farrer et al. 2003; Farrer & Frith 2002; Ruby & Decety 2001).

We are not committed to one of these more specific proposals. We merely mention them in order to make clear that the question of self-other differentiation is mainly problematic for those aiming to explain action understanding solely in terms of mirror neuron processes – like Iacoboni or Gallese. But this is certainly not a position that we wish to defend. In fact, we think the more interesting question for those who subscribe to shared representations is precisely what additional mechanisms are required to explain the differentiation between self and other. This is not only necessary to account for instances of imitative behavior, which is central to most experiments on mirror neurons (cf. Iacoboni 2005), but also action emulation (i.e. achieving the same goal by different means), and cases in which agents actually do the *opposite* of the observed behavior.

Given our current focus on the development of false belief understanding, we are primarily interested in those situations in which infants are sensitive to the reality-incongruent informational states of other agents (i.e. their false beliefs). Instead of opting for 'naked representations', however, we shall follow other researchers (e.g., Leslie et al. 2004, Birch & Bloom 2007; Carlson & Moses 2001) in assuming that infants simply have a tendency to attribute their own representational states to other agents. Thus, we agree with Goldman (2006) that infants' default procedure is to project their own basic concepts onto others.

In section 3, we criticized Baillargeon et al. (2010) for explaining the difference between implicit and explicit false belief understanding in terms of a single inhibition-selection-representation mechanism. More in particular, we suggested that implicit false belief understanding may involve selection processing and response-inhibition as well. This is because infants still have to select a reality-incongruent informational state, and in order to do so they have to inhibit their default tendency to attribute a reality-congruent informational state. In what follows we will explain this proposal in more detail.

Our starting point is Baillargeon et al. (2010)'s notion of a reality-congruent informational state, i.e. an informational state represented on the basis of the infant's own perspective on the scene. We think this notion has potential when it comes to explaining the development of false belief understanding, but it needs to be specified in more detail – in particular in relation to the infant's mode of perspective taking. Furthermore, in line with other dual-system accounts, we postulate two systems in order to account for the development of false belief understanding: (a) an inhibition-selection-representation system, or 'ISRS', and (b) a default attribution mechanism that is responsible for the default attribution of reality-congruent informational states, or 'DAM'. Central to our proposal is the idea that ISRS, like Leslie's Selection Processing system, functions as a 'de-coupling' mechanism. In our model, however, ISRS is not responsible for decoupling *false beliefs*. Rather, it decouples different kinds of reality-congruent informational states.

We propose that there are three ways in which ISRS facilitates the decoupling of reality-congruent informational states, thus allowing infants to understand the reality-incongruent information states of other agents instead. In the first place, ISRS can decouple a reality-congruent informational state by inhibiting its *sensorimotor* properties. The basic idea is that infants by default register reality-congruent informational states on the basis of their own sensorimotor perspective, i.e. their own (intended) movements towards the object. It has been hypothesized that this underlies infants' ability to anticipate the consequences of their own behavior: the brain generates motor-simulations of intended movements by sending efference copies through a forward control mechanism in order to compare them with an ongoing movement to predict its success (e.g., Frith et al. 2000, Blakemore et al. 2002, Blakemore & Frith 2003). In order to understand the goal-directed behavior of *another* agent (i.e. her movement towards the object), three ISRS sub-processes are required: (i) a response inhibition process (infants have to inhibit their own sensorimotor perspective, (ii) a response selection process (infants have to select the sensorimotor perspective of the other agent), and (iii) a representation process (infants have to represent a reality-incongruent informational state that is informed by the other agent's sensorimotor perspective). This first-order mode of decoupling reality-congruent informational states allows us to explain what happens in experiments like the one by Luo & Baillargeon (2005), where infants are able to anticipate the behavior of another agent on the basis of his or her movement towards the object.

One might wonder how the infant is able to select and represent the sensorimotor perspective of another agent. This is precisely where we think the notion of shared representation could play an important role: the infant's perception of the agent's goal-directed behavior in the familiarization trials leads to 'shared resonance', and provides the infant with the sensorimotor information required to execute the observed action. During the test trial, the infant has to select and represent a reality-incongruent informational state on the basis of this sensorimotor information.

Secondly, ISRS can also decouple a reality-congruent informational state by inhibiting the infant's *perceptual* perspective. Our assumption is here that infants by default register reality-congruent informational states on the basis of their *own perception* of the object. In order to anticipate another agents' behavior on the basis of their visual perspective (i.e., what they can or cannot see), the infant has to take offline its own perceptual perspective. This second-order mode of decoupling again involves three ISRS processes: inhibition, selection and representation.

Now we are in the position to explain what happens in implicit false belief experiments, which show that infants are able to anticipate the behavior of another agent on the basis of his or her perception of the object. Consider again the Southgate et al. (2007) experiment, which showed that 25-month-olds correctly anticipated the behavior of a protagonist with a false belief about the location of a ball. According to our model, in order to anticipate the behavior of the agent on the basis of her visual perspective (i.e., what the agent saw in the previous scene) the infant has to inhibit its own perceptual perspective, and represent the perceptual perspective of the other agent instead.

Finally, the ISRS allows infants to decouple a reality-congruent informational state by inhibiting its *cognitive* perspective. Before they acquire linguistic competence, infants are already capable of representing certain proximal goal-directed actions as informational states with a means-end structure. This provides them with a basic understanding of the 'in-order-to' relations that are characteristic for goal-directed behavior, e.g., the agent reaches out in order to grasp the object. What is important about linguistic symbols is that they allow infants to (re)configure informational states in much more complex 'in-order-to' relations, thereby enabling an increasingly sophisticated typing of the distal goal-directed actions of other agents. Our use of the terms 'perceptual perspective' and 'cognitive perspective' is meant to illustrate precisely this difference between perceiving and/or representing proximal versus more distal goal-directed actions.

This last ability is an important requirement for the explicit FBT. For this task requires infants to deal with rather abstract experimental scenarios, i.e. stories or pictures instead of the interacting real-life agents and objects that feature in the implicit FBT. On our view, infants by default register what happens in these scenarios on the basis of their own cognitive perspective. Now in order to verbally predict another agents' behavior on the basis of *their* cognitive perspective, the infant has to take offline its own reality-congruent perspective. This third-order mode of decoupling again involves three ISRS processes: inhibition, selection and representation.

This allows us to explain what happens in the explicit FBT, but it does not yet explain the developmental paradox of false belief understanding, i.e. why infants have more difficulty with the explicit false belief task vis-à-vis its implicit counterpart. Although we cannot offer a concrete solution to this problem here, we do think there a number of options that should be further investigated. In the first place, the explicit false belief task might simply be more difficult because it requires a much stronger form of decoupling. Evidence suggests that explicit false belief understanding indeed places increasing demands on executive functioning. For example, several studies have found robust correlations between explicit

FBT performance and response inhibition (e.g., Perner & Lang 1999, Cole & Mitchell 2000, Carlson & Moses 2001) and working memory (Carlson et al. 2002, Hala et al. 2003, Perner et al. 2002).

However, it seems unlikely that this is the whole story. What studies such as the one by Southgate et al. (2007) show is that especially *verbal interaction* between infant and experimenter crucially contributes to the difficulty of the explicit FBT. Many experiments have found strong correlations between linguistic competence and explicit FBT performance (Dunn et al. 1991, Astington & Jenkins 1999, Gale et al. 1996, De Villiers & De Villiers 2000, Watson et al. 2002, Farrar & Maag 2002). There are several hypotheses about why children have more difficulty with FBTs involving linguistic interaction. Some researchers propose that children need to master its semantics (Moore et al. 1990), whereas others argue that what is required is getting a handle on its syntactic structure (e.g., Hale & Tager-Flusberg 2003, Lohmann & Tomasello 2003). We are not committed to one of these hypotheses in particular, but we like to point out that they are not incompatible with the previous point about the stronger decoupling requirement. It is very well possible that the explicit FBT requires a stronger form of decoupling precisely *because* it involves language. An intriguing possibility is that infants fail the explicit false belief task because there is something that *interferes* with the decoupling process, namely, their verbal interaction with the experimenter. This requires further investigation, however.

6. Closing Comments and Further Research

One of the strengths of the ISRS-DAM model presented in the previous section is that it does justice to the developmental continuity of false belief understanding, and also gives a clear explanation of how the interaction between the two sub-systems provides infants, at each stage of development, with more advanced capacities to understand other agents. In this way, the model is able to avoid two serious problems for dual-system accounts of false belief understanding, namely, (i) how these systems interact and (ii) what this implies for their ontogenetic development (see section 3).

Of course there are several remaining issues that still need to be addressed. One important question has to do with the role of shared representations in the infant's ability to represent and understand another agent's visual and cognitive perspective. A second question concerns the role of inhibition in decoupling reality-congruent informational states. We have argued that ISRS enables infants to inhibit their own reality-congruent perspective in order to represent the reality-incongruent perspective of another agent. But is this decoupling required for all instances in which infants register reality-incongruent informational states? It seems plausible to assume that the amount of inhibition required diminishes as a result of learning, and infants become increasingly skilled at switching between congruent and incongruent perspectives in the long run.

The ISRS-DAM model also offers new directions for future research, for instance with respect to its neurobiological implementation. An interesting idea is to understand DAM, which underlies the default tendency to attribute reality-congruent informational states, as a *simulation* mechanism. Consequently, it could be investigated to which extent DAM recruits the brain areas traditionally associated with the mirror neuron system: the superior temporal sulcus, the inferior frontal cortex, and the rostral part of the inferior parietal lobe (Iacoboni et al. 1999, 2005, Iacoboni & Dapretto 2006, Koski et al. 2002, 2003, Decety et al. 2002, Chaminade et al. 2005). The ISRS processes that enable the representation of reality-incongruent informational states, by contrast, might be facilitated by a 'mentalizing' network, consisting of the anterior cingulated cortex, the temporoparietal junction, the superior temporal sulcus and the temporal poles (Frith & Frith 2003, Amodio & Frith 2006). Further research has to show whether it is possible to establish such a link between DAM and the mirror neuron system, and ISRS and the mentalizing network.

References

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7,* 268–277.

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

Aschersleben, G., Hofer, T., & Jovanovic, B. (2008). The link between infant attention to goal-directed action and later theory of mind abilities. *Developmental Science, 11*, 862–868.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relationship between language and theory-of-mind development. *Developmental Psychology*, *35,* 1311–1320.

Baillargeon, R., Scott, R.M and Zijing, H. (2010). False-belief understanding in infants. *Trends in Cognitive Science 14, 3, 110-118*

Baron-Cohen, S. (1995). Mindblindness: an essay on autism and theory of mind. MIT Press/Bradford Books.

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

Bermúdez, J. (2009). Mindreading in the Animal Kingdom. In: Lurz, R. (ed.) The Philosophy of Animal Minds. Cambridge: Cambridge University Press.

Birch, S.A.J. & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18(5)*, 382-386.

Biro, S., Csibra, G., & Gergely, G. (2007). The role of behavioral cues in understanding goal-directed actions in infancy. *Progress in Brain Research, 164*, 303–322.

Biro, S., & Leslie, A. M. (2007). Infants' perception of goal-directed actions. Development through cues-based bootstrapping. *Developmental Science, 10(3),* 379–398.

Blakemore, S., Wolpert D., & Frith, C. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences, 6,* 237–42.

Blakemore, S. J., & Frith, C. D. (2003). Self-awareness and action. *Current Opinion in Neurobiology, 13(2),* 219–224.

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

Bogdan, R. (2009). *Predicative Minds: The Social Ontogeny of Propositional Thinking*. Cambrdige, MA: MIT Press.

Braver, T.S., Cohen, J.D., & Barch, D.M. (2002). The role of the prefrontal cortex in normal and disordered cognitive control: a cognitive neuroscience perspective*. In: Stuss D.T., Knight R.T., (eds). Principles of frontal lobe function. Oxford: Oxford University Press,* 428–448.

Braver, T.S., & Cohen, J.D. (2001). Working memory, cognitive control, and the prefrontal cortex: computational and empirical studies. *Cognitive Processes, 2,* 25–55.

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

Call, J., & Tomasello, M. (1999). A nonverbal theory of mind test. The performance of children and apes. *Child Development, 70,* 381–395.

Carlson, S., & Moses, L. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72,* 1032–1053.

Carlson, S., Moses, L., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11,* 73–92.

Chaminade, T., Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience, 25* (47), 11055–11059.

Clearfield, M. W., Diedrich, F. J., Smith, L. B., & Thelen, E. (2006). Young infants reach correctly in A not-B tasks: On the development of stability and perseveration. *Infant Behavior & Development*, *29,* 435–444.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9,* 377–397.

Clements, W. A. & Perner, J. (2001). When actions really do speak louder than words? but only implicitly: Young children's understanding of false belief in action. *British Journal of Developmental Psychology*, *19*, 413–432.

Cole, K., & Mitchell, P. (2000). Siblings in the development of executive control and a theory-of-mind. *British Journal of Developmental Psychology, 18,* 279–295.

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. Developmental Science, 1(2), 255–259.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. *In Y. Munakata & M. H. Johnson (eds.), Processes of change in brain and cognitive development. Attention and performance, xxi*. Oxford, UK: Oxford University Press.

Csibra, G., & Southgate, V. (2006). Evidence for infants' understanding of false beliefs should not be dismissed. Response to Ruffman and Perner. *Trends of Cognitive Science 10,* 4-5.

Daprati, E., Frank, N., Georgieff, N., Proust, J., Pacherie, E., & Dalery Jeannerod, M. (1997). Looking for the agent: an investigation into consciousness of action and self consciousness in schizophrenic patients. *Cognition, 65,* 71–86.

Decety, J., Chaminade, T., Grezes, J., & Meltzoff, A. (2002). A PET exploration of the neural mechanisms involved in reciprocal imitation. *NeuroImage, 15*, 265–272.

De Villiers, J., & De Villiers, P. (2000). Linguistic determinism and the understanding of false beliefs. In Mitchell, P., & Riggs, K. J. (Eds.), *Children's reasoning and the mind* (pp. 191–228). Hove, England: Psychology Press.

De Vignemont F. (2004). The co-consciousness hypothesis. *Phenomenology and the Cognitive Sciences 3,* 97-114.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Younblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecendents. *Child development*, *62,* 1352–1366.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 229–240.

Farrer, C., & Frith, C. D. (2002). Experiencing oneself vs. another person as being the cause of an action: the neural correlates of the experience of agency. *Neuroimage, 15,* 596–603.

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: a positron emission tomography study. *Neuroimage, 18,* 324–333.

Farrar, M. J., & Maag, L. (202). Early language development and the emergence of a theory of mind. *First language*, *22,* 197–213.

Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly Journal of Developmental Psychology, 50,* 274–290.

Frith, C., Blakemore, S., & Wolpert, D. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Brain Research Review, 31,* 357–363.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalising. *Philosophical Transactions, Series B, 358 (Special issue on Mechanisms of social interaction (Frith, C. D., & Wolpert, D., Eds.)*, 459–473.

Garnham, W. A., & Perner, J. (2001). When actions really do speak louder than words? but only implicitly: Young children's understanding of false belief in action. British Journal of. Developmental Psychology, 19, 413–432.

Gale, E., deVilliers, P., deVilliers, J., & Pyers, J. (1996). Language and theory of mind in oral deaf children. Paper presented at the *Boston University Conference on Language Development*, Boston, MA.

Gallese, V. (2001). The 'shared manifold' hypothesis - From mirror neurons to empathy. *Journal of Consciousness Studies, 8,* 33-50.

Gallese, V. (2005). 'Being like me': Self-other identity, mirror neurons and empathy*. In Hurley, S. &Chater, N. (Eds.), Perspectives on imitation: From cognitive neuroscience to social science: Mechanisms of imitation and imitation in animals (101–118).* Cambridge, MA: MIT Press

Georgieff, N., & Jeannerod, M. (1998). Beyond consciousnes of external reality: A "Who" system for consciousness and action and self-consciousness. *Consciousness & Cognition, 7,* 465–487.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences, 7,* 287–292.

Gerson, S., & Woodward, A.L. (in press*). Building intentional action knowledge with one's hands.* In S.P. Johnson (Ed.), Neo-constructivism. New York: Oxford University Press.

Goldman, A. (2006). Simulating minds: The philosophy, psychology and neuroscience of mindreading. New York: Oxford University Press.

Gordon, R.M. (1986). Folk psychology as simulation. *Mind & Language, 1,* 158-171.
Grezes, J., Armony, J., Rowe, J. & Passingham, R. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage 18*, 928-937.
Guajardo, J. J., & Woodward, A. L. (2004). Is agency skin-deep? Surface attributes influence infants' sensitivity to goal-directed action. *Infancy, 6,* 361–384.
Hala, S., Hug, S., & Henderson, A. (2003). Executive functioning and false-belief understanding in preschool children: Two tasks are harder than one. *J. Cogn. Dev*., *4,* 275–298.
Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: a training study. *Developmental science*, *6(3),* 346–359.
He, Z., & Baillargeon, R. (2007). Understanding of false belief in 11-month-old infants. *Paper presented at the biennial meeting of the Society for Research in Child Development, Boston, MA.*
Herschbach, M. (2008). Folk psychological and phenomenological accounts of social perception. *Philosophical Explorations*, *11(3),* 223-235.
Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans 42–58.
Hurley, S. (2008). The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral Brain Sciences, 31*, 1–58.
Iacoboni, M. (2005). Understanding others: Imitation, language, empathy. In Hurley, S., & Chater, N. (Eds.), *Perspectives on imitation: From cognitive neuroscience to social science: Mechanisms of imitation and imitation in animals* (pp. 77–99). Cambridge, MA: MIT Press.
Iacoboni, M., Woods, R. P., Brass, M., Hekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*, 2526–2528.
Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience,* 7, 942–951.
Jeannerod, M., & Pacherie, E. (2004). Agency, Simulation and Self-Identification. *Mind & Language, 19,* 113-146.
Kanakogi, Y., & Itakura, S. (2010). The link between perception and action in early infancy: From the viewpoint of the direct-matching hypothesis. *Japanese Psychological Research, 52(2),* 121–131.
Kochanska, G., Coy, K. C., & Murray, K. T. (2001). The development of self-regulation in the first four years of life. *Child Development, 72,* 1091–1111
Kochanska, G., Padavich, D. L., & Koenig, A. L. (1996). Children's narratives about hypothetical moral dilemmas and objective measures of their conscience: Mutual relations and socialization antecedents. *Child Development, 67,* 1420–1436.
Koski, L., Wohlschlager, A., Bekkering, H., Woods, R. P., Dubeau, M. C., Mazziotta, J. C., & Iacoboni, M. (2002). Modulation of motor and premotor activity during imitation of target-directed actions. *Cerebral Cortex, 12,* 847–855.
Koski, L., Iacoboni, M., Dubeau, M. C., Woods, R. P., & Mazziotta, J. C., (2003). Modulation of cortical activity during different imitative behaviors. *Journal of Neurophysiology, 89*, 460–471.
Kovács, A., Teglas, E., & Endress, A (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science, 330,* 1830–1834.
Leslie, A. M. (2000). How to acquire a 'representational theory of mind'. In Sperber, D. (Ed.), *Metarepresentations: A Multidisciplinary perspective* (pp. 197–223). New York: Oxford University Press.
Leslie, A.M., Friedman, O. & German, T.P. (2004) Core mechanisms in "theory of mind". *Trends in Cognitive Sciences 8(12),* 528-533.
Lingnau A., Gesierich B. & Caramazza A. (2009). Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *PNAS 106,* 9925-30.
Lohmann, H. & Tomasello, M. (2003). The role of language in the development in false belief understanding: A training study. *Child Development*, *74,* 1130–1144.
Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychol. Sci., 16,* 601–608.
Luria, A.R. The frontal lobes and the regulation of behavior. In: K.H. Pribram and A.R. Luria, Editors, *Psychophysiology of the frontal lobes*, Academic Press, New York (1973), pp. 3–26
McKay, K.E., J.M. Halperin, S.T. Schwartz, and V. Sharma. 1994. Developmental analysis of three aspects of information processing: Sustained attention, selective attention, and response organization. *Developmental Neuropsychology 10,* 121-132.

Meltzoff, A. N., & Moore, M. K. (1977). Immitation of Facial and Manual Gestures by Human Neonates. *Science, 198,* 75–78.
Meltzoff, A. N., & Moore, M. K. (1994). Imitation, memory, and the representation of persons. *Infant Behavior and Development*, *17,* 83–99.
Meltzoff, A. N., & Brooks, R. (2001). 'Like me' as a building block for understanding other minds: Bodily acts, attention, and intention. In Malle, B. F., Moses, L. J., & Baldwin, D. A. (Eds.), *Intentions and intentionality: foundations of social cognition* (pp. 171–191). Cambridge, MA: MIT Press.
Meltzoff, A. N. (2004). Imitation and other minds: The "like me" hypothesis. In Hurley, S., & Chater, N. (Eds.), Perspectives on Imitation: From Neuroscience to Social Science Vol. II (pp. 55–77). Cambridge, MA: MIT Press.
Meltzoff, A.N. (2006). The "like me" framework for recognizing and becoming an intentional agent. *Acta Psychologica, 124*, 26–43.
Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speakers certainty and uncertainty and its relation to the development of representational theory of mind. *Child Development*, *61,* 722–730.
Moses, L. J., & Flavell, J. H. (1990). Inferring false beliefs from actions and reactions. *Child Development, 61,* 929–945.
Moses, L. J., Carlson, S.M., & Sabbagh, M.A. (2005). On the specificity of the relation between executive function and children's theory of mind. In Schneider, W., Schumann-Hengsteler, R., & Sodian, B. (Eds.), *Young children's cognitive development: Interrelations among executive working memory, verbal ability and theory of mind* (pp. 131–146). Hillsdale, NJ: Erlbaum.
Nichols, S., & Stich, S. (2003). *Mindreading. An integrated Account of Pretence, Self-Awareness, and Understanding of Other minds.* Oxford: Clarendon Press.
Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science, 308(8),* 255–258.
Passler, M.A., Isaac, W. & Hynd, G.W. (1985). Neuropsychological development of behavior attributed to frontal lobe functioning in children. *Developmental Neuropsychology,* 7(2), 131-149
Penn, D. C., Holyoak, K. J., Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav Brain Sci., 31,* 109–178.
Perner J., Leekam S. & Wimmer H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5,* 125–137.
Perner, J. 1991. *Understanding the representational mind.* Cambridge, MA: Bradford Books/MIT-Press.
Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3(9),* 337–344.
Perner, J. & Ruffman, T. (2005) Infants' insight into the mind: How deep? *Science 308*, 214–216
Perner, J., & Clements, W. A. (2000). From an Implicit to an Explicit "Theory of Mind". In Rossetti, Y., & Revonsuo, A. (Eds.), Beyond Dissociation: Interaction between dissociated implicit and explicit processing (pp. 273–294). Amsterdam/Philadelphia: John Benjamins Publishing Co.
Perner, J., Lang, B., & Kloo, D. (2002). Theory of Mind and Self Control: More than a common problem of inhibition. *Child Development, 73,* 752–767.
Perner, J. Leekham, S. & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal Developmental Psychology, 5,* 125–137.
Phillips, A., Wellman, H., & Spelke, E. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition, 85,* 53–78.
Prinz, W. (2002). Experimental approaches to imitation*. In Meltzoff, A. N., & Prinz ,W. (Eds.). The imitative mind: Development, evolution, and brain bases* (pp. 143-162). Cambridge: Cambridge University Press.
Poulin-Dubois, D., Brooker, I., & Chow, V. (2009). The developmental origins of naïve psychology in infancy. In Bauer, P. J. (Ed.), *Advances in child development and behavior* (Vol. 37, pp. 55–104). San Diego, CA: Academic Press.
Press, C., Gillmeister, H., & Heyes, C. (2007). Sensorimotor experience enhances automatic imitation of robotic action. *Proceedings of the Royal Society B: Biological Sciences, 274,* 2639-2644.
Rizzolati, G., & Craighero, L. (2004). The mirror-neuron system*. Annual Review of Neuroscience, 27,* 169–192.
Rizzolatti, G., Fogassi, L., & Gallese, V. (2006). Mirrors in the mind. *Scientific American, 295(5),* 54-61.

Ruby, P., & Decety, J. (2001). Effect of the subjective perspective taking during simulation of action: a PET investigation of agency. *Nature Neuroscience, 4,* 546–50.
Ruffman, T., Garnham, W., Import, A., & Connolly, D. 2001. Does Eye Gaze Indicate Knowledge of False Belief: Charting Transitions in Knowledge. *Journal of Experimental Child Psychology, 80,* 201-224.
Ruffman, T. and Perner, J. (2005) Do infants really understand false belief?: Response to Leslie. *Trends in Cognitive Science, 9*, 462–463
Russell, J. (2007). Controlling core knowledge: conditions for the ascription of intentional states to self and others by children. *Synthese, DOI 10.1007/s11229-007-9203-8.*
Scholl, B. J., & Leslie, A. M. (1999). Modularity, Development and 'Theory of Mind'. *Mind & Language, 14,* 131–153.
Scott, R., Song, H., Baillargeon, R., & Leslie, A. M. (2007). 18.5-month-old infants attribute to others false beliefs about objects' internal properties. *Paper presented at the biennial meeting of the Society for Research in Child Development, Boston, MA.*
Scott, R.M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. Child Development, 80, 1172–1196.
Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology, 18,* 668-671.
Sirois, S., & Jackson, I. (2007). Social cognition in infancy: A critical review of research on higher order abilities. *European Journal of Developmental Psychology, 4,* 46–64.
Sodian, B. (2005). Theory of mind. The case for conceptual development. In Schneider, W., Schumann-Hengsteler, R., & Sodian, B. (Eds.), *Young children's cognitive development. Interrelationships among working memory, theory of mind, and executive functions* (pp. 95–130). Hillsdale, NJ: Erlbaum.
Sodian, B., & Thoermer, C. (2004). Infants' understanding of looking, pointing, and reaching as cues to goal-directed action. *Journal of Cognition & Development, 5(3),* 289-316
Sodian, B., Thoermer, C., & Dietrich, N. (2006). Two- to four-year-old children's differentiation of knowing and guessing in a non-verbal task. *European Journal of Developmental Psychology, 3*, 222– 237.
Sodian et al. (forthcoming)
Sommerville, J. A., Woodward, A., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition, 96,* B1–B11.
Sommerville, J. A., Hildebrand, E. A., & Crane, C. C. (2008). Experience matters: The impact of doing versus watching on infants' subsequent perception of tool use events. *Developmental Psychology, 44,* 1249-1256.
Song, H., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psycholology, 44(6),* 1789–1795.
Southgate, V., Senju, A., & Csibra, G. (2007). Action Anticipation through Attribution of False Belief By Two-Year-Olds. *Psychological Science, 18,* 587–592.
Stich, S. (1978). Beliefs and Subdoxastic States. *Philosophy of Science 45,*499-518
Spaulding, S. 2010. Embodied cognition and mindreading. *Mind & Language, 25(1),* 119-140.
Surian, L., Caldi S, & Sperber, D. 2007. Attribution of beliefs to 13-month-old infants. *Psychological Science*, 18, 580–586.
Tomasello, M., Call, J., Hare, B. (2003). Chimpanzees Understand Psychological States - The Question is Which Ones and to What Extent. *Trends in Cognitive Sciences 7,* 153-6.
Träuble, B., Marinovic, V., & Pauen, S. (2010). Early theory of mind competencies—Do infants understand others' beliefs? *Infancy, 15,* 434–444.
Tsai, C. C., Kuo, W. J., Hung, D. L., & Tzeng, O. J. (2008). Action co-representation is tuned to other humans*. Journal of Cognitive Neuroscience, 20(11),* 2015–2024.
Watson, A., Painter, J., & Bornstein, M. (2002). Longitudinal relations between 2-year-olds' language and 4-year-olds' theory of mind. *Journal of Cognition and Development*, *2,* 449–457.
Wellman, H.M. 1990. *The Child's Theory of Mind*. Cambridge, MA: MIT Press.
Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of Theory of Mind Development: The Truth about False-Belief. *Child Development, 72(3),* 655–684.
Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In Goswami, U. (Ed.), *Handbook of Childhood Cognitive Development* (pp. 167–187). Oxford: Blackwell.
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Woodward, A. L. (1998). Infants Selectively Encode the Goal Object of an Actor's Reach. *Cognition, 69,* 1–34.
Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science, 6*, 297–311.
Woodward, A. L., & Guajardo, J. J. (2002). Infants' understanding of the point gesture as an object-directed action. *Cognitive Development, 17,* 1061–1084.
Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition, 85,* 223–250
Xu, F., Carey, S., & Quint, N. (2004). The emergence of kind-based object individuation in infancy. *Cognitive Psychology, 49*, 155–190.