Mindreading in Infancy

PETER CARRUTHERS

Abstract: Various dichotomies have been proposed to characterize the nature and development of human mindreading capacities, especially in light of recent evidence of mindreading in infants aged 7 to 18 months. This article will examine these suggestions, arguing that none is currently supported by the evidence. Rather, the data support a modular account of the domain-specific component of basic mindreading capacities. This core component is present in infants from a very young age and does not alter fundamentally thereafter. What alters with development are the interactions between core mindreading and other systems, including executive systems, and forms of learning that do not require radical conceptual change.

1. Introduction and Background Assumptions

Prior to 2005 there was a widespread consensus among developmental psychologists that mindreading capacities develop gradually over the first four or more years of life. Admittedly, some believed that mindreading competence is present much earlier (within a few months of birth), but that this is masked by the processing demands of the tasks used by experimenters (Leslie, 1994; Leslie and Polizzi, 1998). Most felt, however, that the extensive evidence collected over the previous two decades, using a wide variety of tasks and methods, was too consistent to be ignored (Wellman *et al.*, 2001).

Then Onishi and Baillargeon (2005) published their groundbreaking paper, using looking time as a measure of false-belief understanding in a simplified task with 15-month-old infants, with positive results. This has been followed by an extensive number of related findings originating from a variety of different labs using a number of different methods. Some use surprise-looking as the dependent measure[1] (Surian *et al.*, 2007; Song *et al.*, 2008; Poulin-Dubois and Chow, 2009;

The views expressed in Sections 3 and 4 of this article were deeply influenced by conversations with Renée Baillargeon, Gergely Csibra, and Victoria Southgate. I am grateful to Ian Apperly, Mark Engelbert, Logan Fletcher, Brendan Ritchie, Josef Perner, Victoria Southgate, and an anonymous referee for their comments on earlier drafts.

Address for correspondence: Department of Philosophy, University of Maryland, College Park, MD 20742, USA.
Email: pcarruth@umd.edu

[1] Not everyone accepts that *surprise* is the appropriate notion to employ in explanation of the infants' looking behavior. But a quarter-century of accumulated experience with the looking-time method means that we can be quite confident that the infants are not, in general, merely responding with interest to low-level novelty in the stimulus. And whether infants form their expectations in advance (issuing in surprise when these are violated), or only realize after the fact that things have not turned out as they should, essentially the same case for early mindreading can be made.

Mind & Language, Vol. 28, No. 2 April 2013, pp. 141–172.

Scott and Baillargeon, 2009; Kovács *et al*., 2010; Scott *et al*., 2010; Träuble *et al*., 2010; Yott and Poulin-Dubois, 2012; Baillargeon *et al*., in press), some use expectancy-looking (Southgate *et al*., 2007; Neumann *et al*., 2009), and some use active helping as the dependent measure (Buttelmann *et al*., 2009; Southgate *et al*., 2010; Knudsen and Liszkowski, 2012). One or two such results could have been dismissed as an aberration, but researchers of all stripes now need to recognize that something real has been discovered (and most do).

This article will discuss the main proposals that have been put forward to accommodate the new data and/or to reconcile it with the traditional developmental account. The first is that what appears to be mindreading competence in infancy is merely implicit in nature (although what is meant by 'implicit' is not always clear, and varies among theorists). This idea will be discussed in Section 2. A second suggestion is that while infants do employ a form of mindreading, it is mindreading of a crude and approximate sort, not involving the attribution of propositional attitudes to other agents. This will be discussed in Section 3. A third proposal is that infant mindreading, although real, bifurcates into two distinct varieties that emerge at different times. (This is just as developmental psychologists had traditionally supposed, only shifted earlier by a couple of years or more.) Stage 1 mindreading enables infants in their first year of life to reason about the desires, perceptual access, and knowledge and ignorance of other agents. Stage 2 mindreading, on the other hand, emerges during the second year of life, and enables infants to represent false beliefs and misleading appearances. This idea will be considered in Section 4. Note that all of these proposals postulate fundamental changes in infants' representational capacities related to mindreading over the first two or more years of life.

In contrast with these accounts, the present article will defend a version of the idea proposed by Leslie (1994; Leslie *et al*., 2004). This is that the domain-specific mechanism implicated in mindreading competence is functional early in infancy. (I shall tentatively suggest: by the middle of the first year of life.) This provides infants with the concepts and core knowledge necessary to represent the mental states of other agents, of all basic types (including beliefs that are false and appearances that are misleading). While the operations of this system probably become more streamlined and efficient with age, its representational capacities do not alter in any fundamental way. Rather, what matures over time are the quality and extent of the interactions between this system and executive, attentional, and planning mechanisms. While this view is by no means conclusively supported by the evidence, it will be shown to accommodate the existing data better than the alternatives.

It is important to note, however, that the proposal of an early-developing core mindreading system is not intended to rule out a subsequent role for learning. On the contrary, it is likely that the system is designed to enrich itself as development proceeds, acquiring new ways of inferring people's mental states from behavioral or contextual cues, for example. There is also likely to be expansion in the range of mental states that the system is capable of representing, adding new non-basic concepts to its stock of representations (such as concepts of socially constructed types of emotion). What is present at the outset is just a capacity to attribute propositional representations of various basic types, including belief.

The view proposed in the present article will take for granted a particular vision of the architecture underlying competence in mindreading tasks. It will be assumed, in particular, that mindreading labor is divided between a domain-specific core system and the agent's own planning systems (in addition to a domain-general working-memory system that is needed for some tasks). This is roughly the architecture outlined and defended at length by Nichols and Stich (2003).[2] The reasons for adopting this assumption will be sketched only briefly here.

How do infants and adults form expectations about what other agents will do? According to one influential view, they adopt an *intentional* or *teleological* stance (Dennett, 1987; Gergely *et al.*, 1995). That is, they attribute a goal to the agent that makes the best sense of her behavior (either present or past); and they assume that the agent will select the most rational and least effortful means of achieving that goal within the constraints provided by the context, and while believing what it would be rational to believe in the circumstances. For example, when familiarized with a geometrical 'agent' who jumps over a barrier to reach another such agent, infants are surprised if the agent doesn't take a direct route to the target once the barrier is removed, but instead jumps in the same trajectory as before (Csibra *et al.*, 2003).

This account is problematic, however. One difficulty is that it involves ascribing to infants a concept of *effort*, as well as theories about what modes of travel are or are not likely to be most effortful for a variety of kinds of agent. (Note that at this age most infants can barely crawl for themselves.) Worse, it involves ascribing to infants a concept of *rationality*. As Nichols and Stich (2003) point out, conceptions of rationality are highly contested even among adults, and it is unlikely that even most adults will have much facility in reasoning with such a concept. But the deepest problem with the teleological account is that it involves ascribing to infants a *theory* of rational planning. As Heal (1996) and many others have argued, this is likely to render mindreading intractable, given that planning can be conducted in indefinitely many contexts in the service of many different goals. Much more plausible is that infants as well as adults should make use of their own planning abilities in figuring out what another agent is likely to do, drawing on their own beliefs for purposes of planning except where these are known to differ from those of the target agent.

What I have just sketched is one of the main arguments underlying simulationist accounts of mindreading. The idea is that, by using one's own planning capacities, one thereby *simulates* the planning of the target agent. Note, however, that nothing in the argument suggests that mindreading competence is simulation-*based*, or that it depends upon introspective access to the results of simulation, as simulation-theorists

[2] I shall remain neutral on the question whether there is a single domain-specific system, or rather a cluster of such systems specialized for specific types of mental state, as Nichols and Stich (2003) maintain.

like Goldman (2006) suppose. (This is all to the good, since there are powerful arguments against simulation-based views; see Nichols and Stich, 2003; Carruthers, 2011.) Rather, what is supported is a mixed account, according to which the domain-specific mindreading system attributes mental states of various kinds to a target agent, while interacting with the subject's own reasoning and planning systems to predict or explain the behavior of the agent (Nichols and Stich, 2003).

What happens when an infant forms an expectation about an agent's behavior, then, will be something like this. On the basis of the previous familiarization trials, the core mindreading system ascribes to the agent the goal of being located next to the other agent (say). Either at the start or end of the experimental trial (depending on whether looking time reflects surprise or retrospective puzzlement), this goal is passed along to the infant's own planning system in the form of the query, 'How does one get *there* [the position of the goal] from *there* [the position of the agent]?' The planning system sets to work to construct a plan, constrained by observable features of the environment, and drawing as needed from among the infant's own currently accessible beliefs. The resulting plan ('Move in a straight line' or 'Move around that obstacle') is attributed to the target agent, thereby forming an expectation about what she should do (either prospective or retrospective). And it is this that is violated if the agent takes a circuitous route in the absence of an obstacle.[3]

As we will see, this sort of mixed domain-specific / domain-general account has the wherewithal to explain much of the data on developmental change, and can do so without postulating fundamental alterations in the representational resources required at different stages of development.

2. Implicit versus Explicit

People commonly use the language of 'implicit' versus 'explicit' to characterize the difference between early competence in nonverbal mindreading tasks and much

[3] An important question concerns the workspace within which these interactions take place. In particular, do they utilize the resources of domain-general working memory, whose contents are generally both conscious and sensory-based (Baddeley, 2006; Carruthers, 2011)? This seems unlikely, at least in the general case. For most instances of simple planning of the kind in question can occur outside of awareness and can operate independently of working memory load. (Think, for example, of maneuvering across a crowded room, or reaching around a glass to pick up a jug of coffee.) And likewise many of the expectations that we form about the behavior of other agents as adults are formed without the recruitment of conscious working-memory processes. Indeed, the phenomenology of much everyday mindreading is that we just *see* someone as being about to act in some specific way in pursuit of a presumed goal, or *hear* the intent behind what they say, without any awareness of how we have arrived at that interpretation. Consistent with these points, concurrent working memory load has little impact on adult performance in first-order false-belief tasks, while having a big impact on performance in second-order tasks (McKinnon and Moscovich, 2007). We will return to the contrast between forms of mindreading that do or do not require working memory in Section 3.

later-emerging competence in verbal ones. But there is no generally-agreed-upon account of what this means. The present section will evaluate some alternatives.

2.1 Explicit States are Available for Report

It is common to identify explicit *tasks* as those that employ some form of verbal or other communicative performance as the dependent measure, whereas implicit tasks are those that use other kinds of measure, such as reaction times. But if one goes on to define explicit *mental states* as those that are manifested in explicit tasks, then all one has really done is re-label the phenomenon one is interested in. We already know that infants show evidence of false-belief understanding in implicit tasks while failing to display such understanding in explicit tasks until some years later. Merely labeling the former understanding 'implicit' and the latter 'explicit' tells us nothing about the kinds of representations that are employed or the computations into which they enter.

2.2 Representing versus Judging

In an early precursor of the recent anticipatory-looking findings, Clements and Perner (1994) found that 3-year-old children in a false-belief task will look toward the correct location when cued by the story narrator, who announces that the story character is about to return to retrieve the desired object and says, 'I wonder where he will look?' Nevertheless, the same children answer incorrectly when *asked* where he will look. Clements and Perner (1994) deem this early understanding *implicit*, by which they mean that the fact of where the character will look is *represented* by the children but not *judged*. On this account, then, infants' understanding of others' false beliefs, and the predictions based upon them, are implicit in the sense that they are computed and represented in the infants' minds, but are not yet believed. Only when they are believed do the children have explicit understanding, which can then guide their answers to verbal questions.

The problem for this account of the infancy data is that it cannot explain the results of the active-helping paradigms (Buttelmann *et al.*, 2009; Southgate *et al.*, 2010; Knudsen and Liszkowski, 2012).[4] This is because the dependent measure in these experiments is an intentional action designed to provide help for the experimenter, in circumstances where what constitutes 'help' depends on drawing inferences from the experimenter's false belief. Southgate *et al.* (2010), for example, first allowed 17-month-old infants to familiarize themselves with two previously-unfamiliar objects. These were then placed into two boxes, and the experimenter left the room. In the interval, a second experimenter appeared and switched the

[4] It would be possible, of course, for a proponent of the represented-but-not-judged view to deny that the data show that the agent's false belief in these more recent experiments is even *represented*, attempting to explain away the evidence in some other terms.

locations of the objects. The first experimenter then returned, pointed at one of the boxes and asked, 'Can you pass me the sefo [a nonsense name]?' (in one condition) or 'Can you pass me it?' (in another). Most of the infants moved toward the other (not indicated) box, interpreting the experimenter's referential intention in the light of her false belief about the locations of the object.

The dependent measure in this experiment appears to be an executively-controlled intentional action. Such actions are paradigmatically informed by the agent's *beliefs*. We have good reason to think, then, not only that the infants *represent* the experimenter's false belief, but also that they judge her to possess it.

2.3 Unconscious versus Conscious

Other theorists have suggested that the knowledge manifested in anticipatory looking, but unavailable for verbal report, is unconscious rather than conscious (Ruffman *et al.*, 2001). There are, however, two broadly different accounts of the functional correlates of consciousness. One is first-order, suggesting that conscious states are those that are widely available to other systems, especially those involved in belief-formation and decision making (Baars, 1988; Tye, 1995). But the infant helping-studies mentioned in Section 2.2 make it unlikely that the infants' judgments about others' false beliefs are unconscious in this sense. For they are available to inform the planning of controlled actions, as well as decisions about which box to approach. The other account is higher-order, maintaining that conscious states are those that one is aware of possessing (Rosenthal, 1986). This is the understanding endorsed by Ruffman *et al.* (2001), who provide evidence based on young children's betting behavior that they are unaware of their own judgments (which nevertheless guide correct anticipatory looking).

Ruffman *et al.*'s (2001) claim may well be correct. Indeed, it seems quite likely that infants and younger children do not yet know that they have knowledge of the false beliefs of a target agent. For such knowledge requires inferential resources equivalent to those needed for success in a second-order false-belief task, where one has to represent that Bill knows that Sally believes (wrongly) that the chocolate is in the cupboard, say. And such tasks are known to be significantly more difficult.

Notice, however, that this account only explains children's comparatively late success in verbal tasks if assertion requires higher-order belief. This seems quite unlikely. Many models of speech production begin from a judgment-to-be-expressed (Levelt, 1989). But none that I know of maintain that speech needs to begin from the higher-order belief that one is making such a judgment. As a result, if expressing a belief that Sally will go to the cupboard only requires a belief with the content, *Sally will go to the cupboard*, then it remains puzzling why younger children should fail to say that she will (indeed, that they should systematically claim the opposite). The fact that they do not have a higher-order belief with the content, *I believe that Sally will go to the cupboard* is of no help. We will return to provide an explanation of the temporal lag between children's performance in nonverbal and verbal tasks in Section 2.6.

Moreover, even if the claim that young children's beliefs are not conscious *could* explain the lag between early competence in non-verbal tasks and late performance in communicative ones, this would fail to establish that the representational resources available to infants differ fundamentally from those of older children. On the contrary, the claim would merely be that infants have beliefs about other's beliefs, but have difficulty forming or reasoning with multiple embeddings of the belief-concept. My goal in this article, however, is to critique accounts that *do* claim that there is a major representational discontinuity in the course of infant development.

2.4 Unstructured versus Structured

In another sense, knowledge is implicit if it is not explicitly represented. Explicit knowledge, in contrast, is knowledge whose content is realized in a symbolic structure of some sort, whose components mirror the conceptual structure of the thought. For example, one can imagine two different ways in which the mindreading faculty might encode the attributional principle, *seeing leads to knowing*. In one, the inference from a representation of the form, JOHN SEES THAT P, to, JOHN KNOWS THAT P, proceeds via the activation of a major premise of the form, SEEING LEADS TO KNOWING. Here the knowledge is explicitly represented. In the other, the system contains a computational rule which generates the conclusion, JOHN KNOWS THAT P, as output when provided with, JOHN SEES THAT P, as input. Here the knowledge that seeing leads to knowing remains implicit in the inference rule in question.[5]

A distinctive feature of explicit forms of representation is their flexibility. Explicit representations are able to combine with many other such representations to issue in new beliefs. For example, when combined with the representation, IF SEEING LEADS TO KNOWING THEN HEARING LEADS TO KNOWING, the representation, SEEING LEADS TO KNOWING can deliver, HEARING LEADS TO KNOWING as a conclusion. The little inference rule envisaged above cannot. But the mindreading knowledge possessed by infants seems to be flexible in this sense. In particular, the ways in which attributed goals and beliefs can interact to issue in expectations of behavior suggests that these states are internally structured. Consider an example from the recent literature.

[5] Radical connectionist theories provide an additional, and very different, way in which knowledge can be merely implicit in this sense. Such theories claim that content is distributed across the nodes in a network in such a way that the conceptual structure of a belief is *not* realized in a structured physical representation. Debates about the computational basis of the mind are ongoing, of course, and cannot be discussed here. However, there seems to be an increasing recognition that connectionist architectures can (and perhaps should) be configured into distinct vectors that realize the conceptual structure of the contents they serve to process (Smolensky and Legendre, 2006). For our purposes it will be enough to show that there is just as much reason to think that the mindreading beliefs of infants are realized in structured states as there is to think that the mindreading beliefs of 4-year-olds are.

Scott and Baillargeon (2009) devised a false-belief task that would involve attributing beliefs about mistaken identity. The infants were familiarized with two 'Penguins', only one of which could split in half like a Russian doll, but which were otherwise identical in appearance. The target agent always initially saw the divisible penguin in its divided state, and used it as a container in which to hide a key. The assumption, then, is that prior to the experimental trial, infants attributed to the agent the goal of placing her key in the divisible penguin. In the false-belief condition the infants watched while a second experimenter assembled the divisible penguin out of its parts, placing it under a transparent cover. The indivisible penguin, meanwhile, was placed under an opaque cover. The agent then appeared, holding the key, and reached either for the transparent or the opaque cover (depending on condition). The infants looked significantly longer when she reached for the transparent cover (while doing the reverse in the true-belief condition). They presumably reasoned that since the agent had always initially seen the divisible penguin in its divided state, she should believe that the complete penguin under the transparent cover was the *in*divisible one, and should therefore reach for the one hidden underneath the opaque cover.

In order to explain the results of this experiment, it seems necessary to attribute to infants thoughts like these: (1) she wants to obtain the divisible penguin; (2) she thinks the penguin under the transparent cover is the indivisible one; hence (given an assumption that the two penguins are always present), (3) she thinks the divisible penguin is under the opaque cover; so (4) she will search under the opaque cover. The inferences involved depend critically on the internal structure of the thoughts ascribed, with differing spatial properties being attributed to distinct objects. In particular, notice that the same individual penguin figures in thoughts (1) and (3), and the same spatial location figures in thoughts (3) and (4).

In addition, of course, if the assumption made in Section 1 is correct, and mindreading competence operates through interactions between a core thought-attribution system and the person's own planning system, then the attitude component in an attributed thought or desire must be distinct and detachable from the content of the attitude. For it is only the latter that is made available to the planning system, either as a goal to be achieved or as an assumption to be relied upon. So I conclude that there are good reasons to think that infant mindreading beliefs are realized in structured representational states.

2.5 The Behavior-Rule Account

In response to Onishi and Baillargeon's (2005) finding of apparent false-belief understanding in 15-month-old infants, Perner and Ruffman (2005) proposed a pair of alternative explanations of the data. One was a low-level associationist explanation, which was soon excluded by the findings of Surian *et al.* (2007) and others. The second was that, instead of representing and reasoning about the false belief of the target agent, the infants might be employing a non-mentalistic behavior-rule such as, 'People will search for an object where it was last in their line of sight.' This, too, would explain the infants' surprise when the agent searches where the object really is, and not where it was last seen.

Perner (2010) argues that behavior-rules of this general sort are capable of explaining all of the evidence of infant mindreading that had emerged up to the time of writing. But he also notes that the use of such rules could be regarded as a form of *implicit* mindreading, since the rules enable infants to track the circumstances in which false beliefs are likely to be produced without explicitly representing them. It will be argued here, however, that the data that have been amassed since 2005 provide significant support for the hypothesis of early mindreading competence, and that the behavior-rule hypothesis is now implausible.[6]

A number of results in the literature might potentially be explained in terms of the behavior-rule mentioned above (Onishi and Baillargeon, 2005; Surian *et al*., 2007; Southgate *et al*., 2007; Buttelmann *et al*., 2009). But there are also many results that it cannot explain. Song and Baillargeon (2008), for example, show that at 14 months infants are surprised when a target agent fails to take account of a misleading perceptual cue when searching for an object. They look longer when an agent who is searching for a blue-haired doll ignores the box that has blue hair sticking out from under the lid (which the infant but not the agent knows to be mere tufts of hair attached to the underside of the lid) and reaches for the other box instead (even though that is where the doll really is). Since the agent in this experiment never saw the doll being placed into either box, the behavior-rule, 'People search where they last saw an object' cannot apply.

There are many other cases, too, where this rule cannot apply. Thus Scott and Baillargeon's (2009) 'Which penguin?' study requires 18-month-old infants to attribute false beliefs about object identity rather than object location, and Song *et al*. (2008) show that at the same age infants seem to understand that an agent's false belief can be corrected by an appropriate, but not an inappropriate, communication. In addition, as we described earlier, Southgate *et al*. (2010) show that at 17 months infants can use an agent's false belief to figure out her referential intentions.

Another behavior-rule that has been proposed to explain the initial results of Onishi and Baillargeon (2005) and Surian *et al*. (2007) is, 'Ignorance leads to error'. (This is not strictly a *behavior*-rule, since ignorance is the absence of a mental state of knowledge. But it only appeals to mental states of a Stage 1 kind, which many think infants can reason about before they are capable of reasoning about false belief.) This would lead infants to expect an agent who is ignorant of the location of a desired object to search in the wrong location, where the object is not. But

[6] Note that behavior-rules might themselves be either implicit or explicit in the sense distinguished in Section 2.4. They might be encoded in some suitably structured explicit representation. Or they might be implicit in a processing procedure linking circumstances with expected behavior. The latter kind of view may be found in the sensorimotor contingency accounts of early mindreading provided by Gallagher (2001, 2004) and Hutto (2004, 2008).

the experiments by Southgate *et al.* (2007) cannot be explained in this way. For the target object was removed from the scene altogether while the agent was absent, rather than moved to another location. Applying the rule 'Ignorance leads to error' in these circumstances should lead the infants to have no expectation about where the agent would search, whereas the experiment actually found that they looked toward the door that she should reach through given her false belief. Moreover, controls to exclude an explanation in terms of such a rule were employed in Scott and Baillargeon (2009) and Scott *et al.* (2010).

It is plain that the existing data cannot be explained in terms of just one or two behavior-rules. On the contrary, many such rules will need to be postulated. It has to be conceded, however, that a determined theorist will always be able to find *some* behavior-rule that could account for any given item of data. Indeed, Povinelli and Vonk (2004) provide a general recipe for finding such rules. They do so in the context of discussing the alleged mindreading capacities of nonhuman primates, but the idea can equally well be applied to the infancy data, as Perner (2010) points out. Their idea is that whenever a mindreading theorist postulates that subjects link a target agent's circumstances to her subsequent behavior via an internally attributed mental state, one could claim that the subjects employ a behavior-rule linking the circumstances to the behavior *directly*, instead.

The resulting claim is that there is *some* (unspecified) set of behavior rules that infants employ. This account makes no predictions of its own, however. Indeed, it is entirely parasitic on positive results provided by the infant-mindreading hypothesis to generate the proposed set of rules. It has to wait for others to provide evidence of infant mindreading, whereupon it constructs a behavior rule to explain the data *post hoc*. It therefore lacks one of the marks of a good empirical theory, which is to be *fruitful*, generating new results and issuing in a progressive scientific research program (in the sense of Lakatos, 1970).

Matters are otherwise for the infant-mindreading hypothesis, of course. It is currently a progressing and increasingly successful research program. It not only predicts that infants should be able to track the goals and beliefs of other agents in simple circumstances, but it enables an open-ended set of further predictions, depending on what else we take infants to believe. Moreover, many such predictions have already been confirmed (Scott *et al.*, 2010; Baillargeon *et al.*, in press). But if it should turn out that these existing studies cannot be replicated, or if additional control experiments provide evidence of non-mentalizing mechanisms underlying the results, then the situation may yet reverse itself. But at present we seem warranted in tentatively endorsing the infant-mindreading hypothesis, based on its record so far.

In addition, it is important to note that the behavior-rule and infant-mindreading hypotheses assume very different explanatory burdens. This is easiest to see if it is assumed that behavior rules need to be learned over the course of the first few months of the infant's life. The challenge, then, is to show, in respect of each proposed rule, that infants have adequate opportunities to learn it in the time available. This seems implausible with respect to a number of the behavior rules that would be needed to explain some of the more recent infancy results.[7] But in any case it is a challenge that behavior-rule theorists have not yet taken up.

The infant-mindreading hypothesis, in contrast, postulates an innately channeled body of core knowledge, or an innately structured processing mechanism (or both), with an internal structure that approximates a simple theory of mind. The explanatory burden, then, is an evolutionary one: it needs to be shown that there were sufficient adaptive pressures among our ancestors for such a mechanism to evolve. There is now an extensive body of work suggesting that this is indeed the case. The gains provided by such a mechanism might derive from enabling so-called 'Machiavellian intelligence' (Byrne and Whiten, 1988, 1997), from facilitating larger group sizes (Dunbar, 1998), from enabling distinctively human forms of cooperation and collaboration (Richerson and Boyd, 2005; Hrdy, 2009), or from any combination of these.

The situation is essentially unchanged if behavior-rule theorists choose to claim that the set of behavior-rules (whatever that should turn out to be) is innate. Since the rules are independent of one another, a separate adaptationist explanation will need to be provided for each. In contrast, since the infant-mindreading hypothesis maintains that what evolved was a system whose internal structure approximates to the structure of the minds of the agents with whom it is to facilitate interactions, the adaptive pressure can derive from social competence quite generally. There is no requirement to find a separate adaptive pressure for each component of the system.

Perner (2010) argues that to discriminate empirically between mindreading and behavior-rule accounts we require experimental data demonstrating competence across varied goals on the part of the agent-to-be-interpreted, combined with a number of distinct belief-inducing cues (such as seeing versus hearing). Yet some of this diversity is already present in the existing data. While most studies involve an agent whose goal is to *locate* a particular object, in Southgate *et al.* (2010) the goal is to be *given* a particular object, and in Scott *et al.* (2010) the goal is to select an object with a given property (rattling when shaken). Moreover, while most studies involve an agent who acquires a false belief from what they have or have not seen, in Baillargeon *et al.* (in press) the belief is acquired by inference from the size of the target object (the toy puppy) and the sizes of the two available containers, and in Träuble *et al.* (2010) one of the true-belief conditions involves beliefs acquired from observing the agent's manual action on the apparatus with her back turned.

While additional variety in the data would be welcome, it already seems reasonable to prefer the hypothesis of infant mindreading to the behavior-rule

[7] For example, in order to explain the results of Scott and Baillargeon's (2009) 'Which penguin?' study, it seems that infants would need to have had opportunities to learn a rule such as this: 'People who have reached for the divisible one of two otherwise similar objects will reach for the location of the hidden member of the pair when the other of the two is visible in its joined state, provided that the construction of that object out of its parts did not take place within the person's line of sight.'

account. Moreover, recall that every determinate rule or set of rules that has been proposed so far has been controlled for in later experiments (and therefore refuted as a general account of the data), and that Povinelli and Vonk's (2004) recipe for generating an *in*determinate set of rules fails to issue in specific predictions.

2.6 Meeting the Challenge: Explaining the Gap

I have suggested that the recent infancy data warrant us in believing that infants in the first half of the second year of life are capable of entertaining structured representations of the beliefs (including false beliefs) of other agents, and of forming simple expectations from these when combined with the infant's own planning abilities (which in turn can draw on her beliefs). But if this is so, the challenge is to explain the gap of two or more years between competence in false-belief reasoning in the second year of life and successful performance in verbal false-belief tasks around the age of four.

How should infant-mindreading theorists respond? Explanations proposed by such theorists prior to the emergence of the new infancy data are no longer applicable. For they were cast in general cognitive terms that aren't specific to language or communication. Thus Leslie and Polizzi (1998) postulated a 'selection processor' that was slow to mature; and Birch and Bloom (2004) appealed to 'the curse of knowledge' combined with immature executive function. But these ought equally to prevent infants from passing nonverbal false-belief tasks as verbal ones.

Since some of the infancy studies are entirely nonverbal, one might wonder whether infants' difficulties in verbal tasks derive from the demands imposed by maintaining a representation of the target agent's beliefs while at the same time processing the questions of the experimenter; for we know that pragmatic components of speech comprehension occupy the resources of the mindreading faculty (Sperber and Wilson, 2002). But there are now a number of successful studies with infants in which speech is employed. Thus both Buttelmann *et al.* (2009) and Southgate *et al.* (2010) employ paradigms in which verbal requests or encouragement are used. In the former the infants are verbally encouraged to help the target agent unlock a box, whereas in the latter they are asked to reach for a designated object (in both cases in conditions of either true or false belief). So it cannot be the mere processing of speech that interferes with the work of the mindreading system in traditional language-based tasks.

Likewise, the same pair of studies seem to rule out another otherwise-plausible hypothesis, which is that the neural connections between crucial components of the core mindreading system (especially the temporo-parietal junction; see Saxe, 2009) and executive systems located in the frontal lobes might not yet have matured sufficiently (Scott and Baillargeon, 2009). One natural idea would be that a thought-attribution system that is frontally-isolated might be capable of driving low-level attentional forms of behavior like looking time while being *in*capable of guiding executively controlled behavior like answering a question.

But the helping paradigms employed by Buttelmann *et al.* (2009) and Southgate *et al.* (2010) make this suggestion difficult to sustain. For in these experiments the dependent measure is request-compliant intentional *action*, which would surely require executive resources. Moreover, the interactions between core mindreading and planning systems described in Section 1 seem also to require the connections between temporo-parietal junction and frontal planning systems to be functioning successfully at these early ages, as well.

The most plausible suggestion is that it is something about language *production* (or the production of communicative actions generally, including pointing gestures, as in Low, 2010) that disrupts successful performance in verbal false-belief tasks (Scott and Baillargeon, 2009). In particular, it seems reasonable to think that the triple burden imposed on the mindreading system in such tasks could prove too much for infants. For in order to succeed in such a task the child needs to do three things: (1) it has to process and keep in mind the mental states attributed to the target agent, as well as have these interact with its own planning systems to generate expectations for behavior; (2) it has to process the speech of the experimenter and figure out the underlying communicative intention; and (3) it has to formulate an action that would serve to communicate the target agent's mental states or likely actions to the experimenter. Since mindreading will be implicated in all three things (Sperber and Wilson, 2002), it makes sense that a capacity to pass verbal false-belief tasks might depend upon maturational expansion of the processing resources available to the mindreading faculty, or on increasing efficiency in the interactions between thought-attribution systems and executive systems, or both.

This suggestion is consistent with recent findings of maturation in the mindreading system. Saxe *et al.* (2009) find evidence of increasing response-selectivity in the right temporo-parietal junction to situations requiring one to attribute thoughts to other agents, which continues through the childhood years (ages 6 through 11). It seems that the region believed to be crucially necessary for thought-attribution becomes increasingly efficient and dedicated in its processing through childhood. Similarly, Sabbagh *et al.* (2009) find evidence that activity in right temporo-parietal junction and dorsal medial prefrontal cortex in 4-year-olds predicts success in a battery of false-belief tasks. These are thought to be two of the crucial components of the mindreading network.

The suggestion is also consistent with the repeated finding that both language ability and executive function correlate with, and are predictive of, performance in verbal false-belief tasks (Astington and Jenkins, 1999; Carlson *et al.*, 2002, 2004; Kloo and Perner, 2003; Milligan *et al.*, 2007; Kovács, 2009). It makes sense that facility with language would reduce the load on components (2) and (3) in a verbal false-belief task. It also makes sense that experience with language (and with communication more generally) might enhance the development of the mindreading system itself, helping to improve its efficiency. In addition one might expect that better executive function abilities (and especially abilities to inhibit or suppress a salient representation, as we will see in Section 4) would enhance the performance of component (1) in a verbal false-belief task. Moreover, the problem of juggling the demands of the three components, dividing attention and switching appropriately between them, is itself a problem for executive function.

While the idea remains to be directly tested, it seems reasonable to conclude from this discussion that it is the triple processing demands imposed by verbal false-belief tasks that makes them especially difficult for younger children.

2.7 Puzzles for the Account

One puzzle for the account sketched in Section 2.6 is why previous findings of false-belief understanding using anticipatory looking should have failed to find such evidence for children younger than three (Clements and Perner, 1994; Ruffman *et al*., 2001). For in these experiments no verbal response was required of the children, any more than in Southgate *et al*. (2007) and Neumann *et al*. (2009). The obvious difference, however, is that in the latter studies the scenarios unfold completely nonverbally, whereas in the former there is an accompanying verbal narrative. In addition, in the latter the signal for anticipatory looking is itself nonverbal, whereas in the former there is a verbal prompt ('I wonder where he will look'). Since processing of the pragmatic component of speech is known to occupy the resources of the mindreading faculty, it makes sense that the processing demands of these previous studies might have outstripped the abilities of younger children.

If this response is on the right lines, however, then that gives rise to another puzzle: how is it that the active-helping studies of Buttelmann *et al*. (2009) and Southgate *et al*. (2010) produced positive results with 18-month-old children, given that both studies involved experimenter speech? A partial answer may be that these studies involved much *less* speech overall than did those of Clements and Perner (1994) and Ruffman *et al*. (2001), in which there was a continuous experimenter narrative throughout. In addition, Southgate *et al*. (2010) provided the infants with additional memory support in the final phase of the experiment: the two boxes were opened so the child (but not the experimenter) could see which item was in each box. This meant that the child did not need to remember the location of the items, but merely had to recall that they had been switched in the experimenter's absence. Moreover, in Buttlemann *et al*. (2009) most of the infants helped the experimenter without requiring a verbal prompt (which was only given if five seconds had elapsed without helping); and for those who were given a prompt, this took the form of encouragement to help, not encouragement to make a prediction. So the load placed on the mindreading system may have been less.

Another concern with the account provided in Section 2.6 is that there have been completely nonverbal false-belief tasks that much older children have failed. In particular, Call and Tomasello (1999) devised a test that could be used with 4 and 5-year-old children as well as with apes. Many of the children failed, with failure closely associated with failure in verbal false-belief tasks. The structure of the task in question was much more abstract than usual, however. In particular, in the false-belief condition subjects had to reason that the target agent had *some* false belief (without knowing *what* they believed), drawing implications from that.

The experiment involved a hider (who placed a target object in one of two boxes out of the child's sight) and a communicator, who gave the children a visual clue by placing a sticker on one of the boxes.[8] In the true belief condition the communicator observed the hider position the target and then (after leaving the room briefly) informed the child by placing the sticker. (The child's task was to retrieve the object.) In the false-belief condition the communicator watched the initial placement of the object, then left the room while the hider switched the positions of the two boxes. (Note that the child could therefore see that the location of the object had been switched without knowing *which* box it was located in.) When the communicator returned, the test was to see whether the children selected the opposite box from the one the communicator indicated (as they should if they understood that he had a false belief about the location). In this task the child has to represent that the communicator has *some* false belief about the location of the object (without knowing what that belief is, since the child is ignorant of the initial location of the object before the communicator left the room), whereas in the tasks that infants pass, they just need to ascribe a specific belief to the agent (which happens to be false, without needing to represent explicitly *that* it is false, as we will see in Section 4).

A final puzzle concerns the false-belief experiment conducted with 18-month-olds by Knudsen and Liszkowski (2012). For the dependent measure in this study was a communicative point. If it is the demands of communicative action that explain 3-year-olds' failure in verbal false-belief tasks, then how is it that these infants succeeded? In answer, recall that the account sketched in Section 2.6 involves a combination of three distinct loads placed on mindreading: (1) mental states of an appropriate sort need to be attributed to the target agent, recalled, and used for simulative planning; (2) the communicative intent behind the experimenter's speech needs to be interpreted; and (3) a communicative action designed to have some specific effect on the mind of the experimenter needs to be formulated. Although Knudsen and Liszkowski's study involved component (3), there was very little speech in the course of the experiment, and none that needed to be interpreted to give rise to an appropriate communicative goal (which was to prevent the experimenter from coming into contact with a disgusting object).

The triple-load account of infants' failure on verbal false-belief tasks suggests that they fail because a combination of multiple processing demands overwhelms the resources of their mindreading systems. But can it also explain why young children don't just become confused, and answer at chance, but give answers that are systematically incorrect? In general, wrong answers are a reality-based response.

[8] Note that this is a communicative action, which would therefore engage the resources of the mindreading system. But this fact by itself cannot explain children's failure in this task, since younger children pass the minimally-verbal tasks we have already discussed.

But this is what we should expect if any one of the three components collapses under load. If a representation of the false belief of the agent is lost, then the language production process is likely to default to the next-most-relevant answer, which is what the subjects themselves would believe or do. If the experimenter's question is incorrectly interpreted, the next-most-likely interpretation will be a question about what is really the case, or what really should be done. And if the production process goes awry, again the most likely error will be to select the infant's own belief or own likely action as the target for report.

Overall, then, it appears that the triple-load account outlined in Section 2.6 has the resources to handle all or most of the existing results in the literature in a satisfying way. It should also be testable. For it predicts that one should be able to make false-belief tasks harder or easier for infants by ratcheting up or down the processing demands of any one of the three components. Consistent with this suggestion, Buttelmann *et al*. (2009) found that 16-month-olds failed the same tasks that 18-month-olds passed, although 16-month-old infants have passed other kinds of false-belief task. The difference may be that the experiment required some processing of experimenter speech in addition to tracking and reasoning about the beliefs of the target agent.

2.8 Conclusion

The hypothesis of a form of early mindreading concerning other people's thoughts is well supported by the data that has accumulated since 2005. Although the hypothesis is not *proven* by the data (no hypothesis ever is), it is significantly more plausible than any competing implicit-knowledge approach. So we can tentatively conclude that in some way, or to some extent, infants in the first half of the second year of life have explicit (structured) representations of the beliefs and false beliefs of other agents. The question is in what way, and to what extent. This is where we go next.

3. Propositional versus Non-Propositional

In considering the contrast between infant mindreading and the capacities displayed by older children and adults, some have argued that infant mindreading is non-propositional whereas the mindreading of older children is fully propositional (Apperly and Butterfill, 2009; Apperly, 2011). This account, too, postulates a fundamental dichotomy in mindreading representational resources between infancy and childhood.

3.1 Non-Propositional Mindreading

We noted in Section 1 that the domain-specific component of the early mindreading system is likely to undergo significant enrichment with development, adding new mental-state concepts alongside its stock of conceptual primitives, and acquiring information about new ways in which mental states can get formed. But Apperly (2011) goes much further, arguing that there is a representational discontinuity between a swift and efficient initial system, which remains intact largely unchanged in adulthood, and a slower developing and slower operating, but nevertheless highly flexible, mindreading system.

Apperly (2011) draws an analogy between infant mindreading and infant mathematics. We know that infants (as well as adults) have a system for estimating the numerosity of a set (Barth *et al.*, 2006). But this system is incapable of assigning a precise number to the set. Instead it delivers approximate representations of number, which one might translate into English using locutions such as, 'It has around 20 members' (except, of course, that the infants' representation does not embed a precise representation of the number 20). It is only when children acquire language, and especially when they learn to use numerals for counting, that they form a conception of precise numerical quantities. Similarly, Apperly thinks, infants initially have only crude and imprecise representations of the beliefs and false beliefs of other agents, which fail to discriminate between the many different ways that the agent might be representing the state of affairs in question. Only as they acquire language, and become capable of more advanced forms of executive function and perspective taking, do children acquire a conception of belief as a relation between a subject and a fine-grained proposition.

Apperly (2011) presents evidence for his view using priming experiments that are designed to tease apart the properties of the initial modular system from the later-developing one. He shows that adults automatically compute the visual perspective of another, and represent what can be seen from that perspective, but without representing *how* it will be seen. Adults whose task is to estimate the number of dots on the walls of a room containing another agent, for example, are slower and less accurate when the visual perspective of the other agent differs from their own. This suggests that two conflicting number representations are produced, one of which is the number of dots as seen by the subject and the other of which is the number of dots as seen by the other agent. This latter representation is automatically computed by the mindreading system, even though it is irrelevant to the task demands, and even though it actually interferes with that task.

Apperly's (2011) reason for thinking that the mindreading system is not computing a proposition, or a fine-grained representation of the way in which the scene is represented by the other agent, is that no interference effects of the above sort occur with such properties. For example, adults are no slower to report a numeral displayed on the desk in front of another agent when that numeral will appear differently to the other agent than they are when it will appear the same. Thus if the subject and other agent have opposite perspectives, there is no difference in reaction times between a case in which an '8' is displayed (which will appear the same to them both) and a case in which a '9' is displayed (which will appear as '9' to one of them and as '6' to the other).

Apperly's reasoning is problematic, however. To see this, notice that Level 2 perspective taking (of the sort required for a difference in reaction times to emerge in this task) is quite demanding of the resources of general-purpose sensory-based working memory, and is by no means a purely mindreading matter. Nor does it seem to have anything to do with the kind of content that is represented (propositional versus non-propositional). In order to figure out that a person sitting opposite will see the numeral '9' as '6' one has to start from one's visual representation of the symbol and flip or rotate its visual image through 180 degrees, noting the result. This is a demanding task, and not the sort of thing that a mindreading mechanism could be expected to perform alone. On the contrary, it will involve a complex interplay between mindreading, executive systems, and the visual system to perform such a task. And by the same token, success in such a task doesn't demonstrate a novel way of representing the contents of other people's thoughts, but just more advanced interactions between the mindreading system and others.[9]

Apperly (2011) thinks, quite reasonably, that the operation of the initial core mindreading system is likely to be automatic. In contrast, Back and Apperly (2010) argue that other people's beliefs (whether true or false) are *not* automatically computed independently of task demands. We thus have reason to think that the initial mindreading system does not represent beliefs. Back and Apperly (2010) required adults to view sequences of photographs involving two human actors, who sat on opposite sides of a desk. The male actor manipulated the location of an object, placing it first under one cup and then under another. Subjects were given the task of keeping track of the location of the object, answering a question about its location at the end of the slide-sequence. In some conditions the female actor left the room while the male actor made a further change in the location of the object (thereby causing her to have a false belief).

On some trials subjects were probed in the middle of the slide-sequence with a nonverbal question, either about the woman's belief or about the current location of the object. (In the first case a picture of the woman and a 'thought bubble' depicting a location of the object was used; in the second case a picture of the two cups and the object located in one of them was used; in both cases preceded by a question mark.) Subjects' task on these trials was to press a key to answer either 'Yes' or 'No' to the implied question as quickly as possible, and their reaction-time was measured. Note that no predictions need to be made regarding the likely behavior of the female agent. Subjects simply had to report what she believed. Since *representing* her beliefs ought to be a function of the core

[9] A similar deflationary explanation can be provided for the results of Low and Watts (forthcoming), who suggest that the early mindreading system is unable to represent beliefs about object identity. Their task, too, is demanding of the resources of working memory and would have required the rotation of a visual image through 180 degrees.

mindreading system (on my account), the use of executive resources should not be required.

What Back and Apperly (2010) found was that subjects were slower to respond to belief probes than to reality probes, irrespective of whether the female actor's belief was true or false. Subjects also made more errors in the former condition than in the latter. From this they conclude that subjects were not automatically computing the beliefs of the agents in the slide-sequence, but were only doing so retrospectively, in response to the probe. But notice that although the probe cue was pictorial (aside from the accompanying question mark), it nevertheless needed to be interpreted. The very act of interpreting the probe and formulating a response would therefore have engaged the resources of the mindreading system. If so, then we can predict that such tasks will be more demanding, because placing a double load on mindreading (tracking the woman's beliefs and interpreting the probe). They should therefore be slower to execute, and should also be more prone to error when done under time pressure.

Thus even if we suppose that the adults in these studies had automatically computed the woman's beliefs, we can still explain why they should be slower to respond to a belief-probe than to a reality-probe, and why they should make more errors in doing so. For the former will require them to allocate mindreading resources to sustain the belief-representation in place while they process the intent behind the visually-based question and select the correct response. In response to a reality-probe, in contrast, only the interpretive component of the task will require the resources of the mindreading system. For the true location of the object is represented in parallel elsewhere in the cognitive system.[10]

Back and Apperly's (2010) results can therefore be explained, and explained consistently with claiming that representations of others' beliefs are often computed automatically, independently of task requirements.

3.2 Dual Systems of Mindreading

Apperly's (2011) main motivation is to defend a dual-systems view of human mindreading competence. One of these systems is fast and inflexible and the other is slow and flexible. The suggestion about the differing representational resources of the two systems is supposed to aid in this endeavor. For Apperly thinks that ascribing propositional representations to other agents is likely to be both computationally demanding and highly unencapsulated, potentially drawing

[10] What of Back and Apperly's (2010) finding that when adults are instructed to pay attention to the woman's beliefs during the slide-sequences the delay in responding disappears? Does this show that only in this condition are the woman's beliefs represented? It does not. For one might suppose that in these circumstances subjects will have pre-prepared an answer to an anticipated question about her beliefs, or about the object's location (which they were still required to track).

on any of the mindreader's background beliefs. But it is difficult to see why these claims should be true.

Consider a false-belief task of the sort presented to infants: a doll that an agent has been playing with is first placed in a blue box and then, while the agent is absent, is moved to a green box. There seems little reason to doubt that the infant itself initially has beliefs with the content, *the doll is in the blue box* while representing the presence of the target agent, and then later, *the doll is in the green box* while noting that the agent is not present when the doll is moved. In order to use these propositions to generate a propositional representation of the target's attitude, it is quite straightforward what happens: during the initial sequence the infant infers from the propositions it has represented, *the agent thinks: the doll is in the blue box*, relying on the attributional principle, *seeing leads to believing*, or some-such. It then does not update this representation when the doll is moved. In a true-belief condition, in contrast, when the doll is moved in the presence of the agent the infant just needs to update its representation of the agent's belief, now judging the content, *the agent thinks: the doll is in the green box*. There seems no reason why these propositional representations could not be formed swiftly and efficiently.

What is surely likely to be true, however, is that infants will have only a limited number of cues that they can use for ascribing thoughts to other agents, at least at the outset. (These will thereafter by enriched by learning, as noted in Section 1.) And they are also likely to be fairly undiscriminating in the propositions that they attribute to other agents. Rather, the main determining factor will be salience *to the infant*. It is the propositional representation of the event in question that the *infant* encodes (or the most salient among such representations) that will be selected to embed within the scope of a belief-ascription. For example, if the infant has itself generated more than one description (perhaps thinking both, *the doll is in the blue box*, and, *the doll is in the box on my left*) then it presumably makes its selection at random, or in terms of salience or some other such property, when formulating what it is that the agent thinks. For at this stage, one might conjecture, an infant will fail to appreciate that the agent is less likely to form a belief with the content, *the doll is in the box on the infant's left*. But there is nothing here to suggest anything deep about the infant's conception of belief in comparison with the adult one. And in many circumstances it will not lead to any error, since what is salient to one person is likely to be salient to another.

I should emphasize that I do not wish to deny the existence of dual systems involved in mindreading, however. For the idea of dual systems for reasoning is now widely accepted in cognitive science (Evans and Over, 1996; Sloman, 1996, 2002; Stanovich, 1999; Wilson *et al*., 2000; Kahneman and Frederick, 2002; Kahneman, 2011). Although terminology has differed, many now use the labels 'System 1' and 'System 2' to mark the intended distinction. System 1 is supposed to be fast and unconscious in its operations, issuing in intuitively compelling answers to reasoning problems in ways that subjects themselves have no access to. System 2, in contrast, is supposed to be slow and conscious in its operations, and is engaged whenever we are induced to tackle reasoning tasks in a reflective manner. Many theorists now accept that System 1 is really a *set* of systems, arranged in parallel, while believing that System 2 is a single serially-operating ability. Presumably, a core mindreading system would be included among the former.

While some scientists have probably thought of the two systems as being wholly distinct, existing alongside one another in the human mind, such an idea has come under increasing pressure (Frankish, 2004, 2009; Carruthers, 2006, 2009). Rather, many in the field have come to accept that the defining feature of System 2 is just that it makes use of the domain-general working memory system, whereas System 1 systems do not (Stanovich and West, 2000; Barrett *et al*., 2004; Evans, 2008; Stanovich, 2009). For example, System 2 processes tend to collapse under concurrent working memory load, whereas System 1 processes don't (De Neys, 2006).

On this account, System 2 processes can be indefinitely flexible because when one operates in reflective mode one can ask oneself questions, search for related memories, constrain verbal inferences to accord with one's explicit beliefs about standards of good inference, and so on. In the domain of mindreading, this is the sort of reasoning one engages in when playing detective, or when puzzling about the reasons for a colleague's outburst. By its nature it is slow, and in principle any of one's beliefs can be brought to bear to have an impact on the result.

None of this shows that there are two mindreading systems, however. Rather, a mindreading system of some sort will be one of the System 1 systems, and System 2 is a general-purpose system, dependent on the controlled use of working memory as well as explicit beliefs and the outputs of System 1 systems. And while one might have many beliefs relevant to mindreading that can become active in System 2 mode (either in general, or in some specific mindreading context), these will fail to constitute a mindreading *system*, let alone one with a radically distinct set of representational resources.

Moreover, it is important not to conflate the idea that there is a distinctive mode of mindreading that depends upon working memory, on the one hand, with the claim that increased working memory abilities are what explain the transition between early failure and later success in verbal mindreading tasks, on the other. For while children's working memory abilities do indeed correlate with their success in verbal mindreading tasks (Keenan, 1998), on some analyses they make no contribution to mindreading independent of the contribution they make to executive function abilities (Carlson *et al*., 2002, 2004), and on others they make no contribution to improvements in mindreading ability over a 6-month period (Slade and Ruffman, 2005).

Working memory will often be important for understanding and remembering the stimulus materials presented in an experiment, of course, but this does not mean that it makes an important contribution to what one does with those materials thereafter. In many cases it seems that thought attributions combined with simple forms of planning (which do not themselves require the resources of domain-general working memory) can take place successfully without it. For as we noted in Section 1, first-order false-belief performance is only marginally impacted by working memory load (McKinnon and Moscovitch, 2007), whereas forms of mindreading that require Level 2 perspective taking (Lin *et al.*, 2010) or reasoning about multiple embeddings of belief (McKinnon and Moscovitch, 2007) collapse under such load.[11]

3.3 Conclusion

There seems no reason to think that the early mindreading system is incapable of attributing propositional thoughts to other agents. On the contrary, since infants *have* propositional thoughts from the outset themselves (as Apperly, 2011, acknowledges), they can take whatever proposition they have used to conceptualize the situation seen by the target agent and embed that proposition into the scope of a 'thinks that' operator. Moreover, although mindreading, like any other form of human reasoning, can be conducted either intuitively or reflectively, there is no reason to think that the representational resources available in the two modes are fundamentally distinct.

4. Two Stages versus One

Most developmental psychologists (whatever their other differences) are united in thinking that the representational resources available for mindreading develop in (at least) two stages (Wellman, 1990; Baron-Cohen, 1995; Gopnik and Meltzoff, 1997; Nichols and Stich, 2003; Scott and Baillargeon, 2009).[12] The first of these involves the notions of desire, perceptual access, and knowledge and ignorance, and the second expands on this to include an understanding of the representational nature of the mind (resulting in a capacity to attribute misleading appearances and false beliefs). For our purposes, however, the version of this hypothesis to be considered is that Stage 1 mindreading precedes Stage 2 by at most a year. For in light of the arguments of Sections 2 and 3, it is reasonable to assume the presence of Stage 2 mindreading competence by the middle of the second year of life, as manifested in a range of nonverbal tasks.

While the developmental difference adverted to here is real, I shall argue that it doesn't map onto two distinct mindreading systems. Nor are there any

[11] Is it a problem for my account that performance in first-order false-belief tasks is impacted *at all* by working memory load? I suggest not. For this seems readily explicable in terms of a shared need for attentional resources. Mindreading performances of all sorts are likely to depend on appropriate allocations of attention. Any task that divides attention will thus be apt to have some impact on performance, albeit a minor one.

[12] One notable exception to this generalization is Leslie (1994; Leslie *et al.*, 2004), who has always defended a modular one-stage account of the sort being proposed here.

deep disparities in the conceptual resources involved. Rather, the difference arises because of early limits on the interaction between the core mindreading system and executive systems.

4.1 Mindreading and Executive Control

Recall from Section 1 the assumption that mindreading involves interactions between a core system and the agent's own planning mechanisms. Then consider, in this light, the difference between ignorance-involving mindreading predictions and those that depend on attributions of false belief. Consider the former first. In order to predict what someone with a given goal will do, who is ignorant of some fact, the mindreading system will need to ask the planning system to figure out how to achieve the goal *without relying on the fact in question*. Since that fact is likely to be highly salient in the context, this is then a challenge for executive function: to 'hold down' or suppress the representation of the fact, insuring that the planning system doesn't rely on it in the course of its reasoning.

Now the important point is this: in order to predict what someone with a given goal is likely to do who has a *false belief* about something, the task for executive systems becomes significantly harder. For the mindreading subject knows the true state of affairs. Hence this representation will need to be suppressed for planning purposes, much as representations of a fact that an agent is ignorant of need to be suppressed. But in addition, a representation of what the target agent believes to be the case will need to be sustained and held in place for planning systems to utilize when they do their work. So not only does a representation need to be suppressed, but an alternative representation needs to be actively maintained. Naturally, this will place significant additional demands on executive function.

It is only to be expected, therefore, that infants should pass ignorance-tasks before they pass false-belief tasks. (Similar points apply to misleading-appearance tasks.) But the limitation need not be a limitation on the representational resources of the core mindreading system. We don't need to postulate a pair of distinct mechanisms or distinct tacitly held theories in order to explain the data. Rather, all of the representational resources required for processing false beliefs and misleading appearances can be present at the earlier stage, and it is the same planning systems that are implicated in both. But these resources are prevented from manifesting themselves in the younger infants' performance because executive functioning is not yet up to the task.

Note that on the account sketched here and in Section 3.2 the distinction between true belief and false belief can be left implicit in the infant's procedures for updating belief attributions. The difference between true belief and ignorance is the difference between someone who does, or does not, think that the doll is the blue box (for example), depending on whether she was or was not present when it was placed there. If the doll is thereafter moved to the green box in the presence of the agent then the infant updates the representation of what the agent thinks. But if the agent is absent at this time, the infant does not update, but maintains the representation, *the agent thinks: the doll is in the blue box*. This is a false belief. But the infant does not need to represent it as such.[13]

Someone might seize upon the point just made to insist that there is a radical shift in representational resources between Stage 1 and Stage 2 mindreading, after all. It will be a shift between representing beliefs (which happen to be either true or false) and representing beliefs *as* true or *as* false. But in fact the latter does not entail a transition to a 'representational theory of mind', or anything of the sort. Rather, the conceptual resources needed to formulate explicit concepts of true and false belief are likely to be present at the earlier stage also. For a true belief that *P* can be defined as a conjunction of a belief that *P* with *P*, whereas a false belief that *P* can be defined as a conjunction of a belief that *P* with *not P*. The need for such concepts may only become salient to children somewhat later (perhaps when they begin to employ and interpret language). But the formation of an explicit concept of false belief can count as a conservative extension of conceptual resources that were available previously, rather than a radical conceptual change.

It is a prediction of the account sketched here that if false-belief tasks can be devised that either reduce or don't require the resources of executive function, then false-belief understanding should manifest itself in infants at the same age that they display competence in desire–ignorance reasoning. Kovács *et al.* (2010) provide evidence that can be interpreted in just this light. I shall first discuss the experiments that Kovács and colleagues conducted with adults. This will set the stage for their studies with 7-month-old infants. For what the adult experiments appear to show is that the mindreading system automatically computes the beliefs of other agents even when irrelevant to task demands, and without utilizing executive function abilities.

[13] Seen from this perspective, it is doubtful whether some recent false-belief tasks really require the attribution of a false belief to the agent. In the rattling cups task, for example (Scott *et al.*, 2010), once the infant suppresses the fact that the spotted cup rattles (because the target agent is ignorant of this fact), the infant's own in initial intuition that the two similar striped cups are likely to share other properties remains in place, and can be accessed by the infant's own planning system when figuring out what the agent should do. Likewise in the collapsible toy-puppy task (Baillargeon *et al.*, in press), once the fact that the toy collapses like a concertina is suppressed (because the target agent did not observe it), the infant's own belief about the relative sizes of the puppy and the small container can drive the simulated planning, resulting in an expectation that the agent will search in the large container. The belief that the puppy cannot fit into the small container does not need to be attributed to the agent at any stage. The same is *not* true, however, of other tasks, including change of location tasks. For here the infant no longer has a belief about the location of the target that can be accessed for planning purposes when simulating the target agent. Rather, an explicit representation of the form, HE THINKS IT IS IN THE BLUE BOX (where it was previously), will need to be employed. Notice, then, that the perspective advanced here undermines the probative force of some false-belief tasks. But at the same time it reduces the evidential burden required, by demonstrating that the representational resources required for a false-belief task are no different from those required for a true-belief task or an ignorance task. The difference lies rather in the procedures for attributing (and updating or failing to update) thoughts to another person.

4.2 Automatic False Belief Representation in Adults

Kovács *et al*. (2010) presented adults with a simple reaction-time task in which mindreading would play no direct part. Subjects watched as a variety of animated scenarios unfolded. In all of them a ball disappeared behind a screen, reappeared, and then either returned behind the screen or left the stage entirely. The screen then dropped, and the subjects' task was simply to press a button as quickly as possible if the ball was there. In some cases it was, even though it had been seen leaving the stage previously; in some cases it was not, even though it had been seen returning behind the screen and should therefore have been present. Not surprisingly, subjects' reaction times were quicker in cases where they held a true belief about the presence of the ball than in cases where they falsely believed that the ball would not be present.

In some of the trials another agent was visible in the video, who also observed some or all of the movements of the ball. The subjects' attention was not drawn to the presence of the observer, who played no part in the on-going activity. But sometimes the observer left during the movements of the ball, in the critical trials failing to observe that the ball had returned to the stage and had disappeared behind the screen. The observer would therefore have had a false belief that there was no ball behind the screen. What the experimenters found is that subjects' reaction times were slowed just as much in this condition as in the condition where they themselves held a false belief that there was no ball behind the screen. It seems that subjects were automatically computing the beliefs, and false beliefs, of the observer, and that these third-person expectations were priming (or in this case inhibiting) the subjects' own action systems just as their own expectations did. This finding with adults has now been replicated by a number of other labs (Ian Apperly, personal communication).[14]

Strictly speaking, all that the data show is that people represent the content of the incidental agent's belief, not that they attribute a belief with that content *to* the agent. It would be sufficient to slow down people's reactions if the representation, THE BALL IS NOT THERE were contained within the mind (caused by the fact that the ball was not there when the incidental agent left the scene), even if there were no representation, HE THINKS: THE BALL IS NOT THERE. But if mindreading operates like all other conceptual consumer systems for perceptual output, this suggestion is unlikely. The normal case is for conceptual contents produced by automatic processing of perception to be bound into the contents of the resulting perceptual states, indexed to the individual or event that caused them (Kosslyn, 1994), or contained in an object-file linked to the thing in question (Pylyshyn, 2003).

[14] Note that a crucial difference between these experiments and those of Back and Apperly (2010) discussed in Section 3.1, is that no concurrent load was placed on the mindreading faculty. Subjects did not need to interpret a cue or answer a question. They simply had to press the button as fast as they could if the ball was visible when the barrier dropped.

4.3 False Belief at Seven Months

Kovács *et al*. (2010) presented infants with exactly the same videos that had been employed with adults, only in this case using looking time as a measure rather than reaction time. They showed infants the same animated sequences in which a ball moved behind a screen, emerged again, and either left the stage or returned behind the screen. Some of these sequences took place while an incidental agent was present in the video, and in some cases the agent left the stage before the ball reappeared from behind the screen and left (thereby giving rise to a false belief about the location of the ball). The experimenters found that the infants were 'primed' to look longer when the screen dropped to reveal the absence of a ball (thereby violating the agent's false expectations, and despite the fact that the infants themselves knew that the ball would be absent), just as adults in similar circumstances were slower to respond with the judgment that the ball is present when it violated another agent's false expectation.

In this looking-time task the infants don't need to make a prediction. All they need to do is compute that the other agent, who was absent when the crucial 'leaving from behind the screen' event occurred, will expect (falsely) that the object is behind the screen. When the screen is lowered to reveal that the object is not there, the presence of a violated expectation attributed to another agent makes this event seem interesting (and makes the infant look longer), even though the infant itself expected the object to be absent.[15]

This is a task with no executive demands beyond those involved in attending to the unfolding events. As such, it is a 'pure' mindreading task. That infants pass it at the age of seven months suggests that representations of false belief are automatically computed by the mindreading system as soon as that system begins to function at all. (Moreover, we have seen in Section 4.1 that it is just as easy to represent false beliefs as it is to represent true ones.) It is reasonable to conclude that there is just a single core mindreading system that is available early in infancy. This contains the resources to represent beliefs that are false and appearances that are misleading in addition to goals, perceptual access, and knowledge and ignorance. The behavioral distinction between 'Stage 1' and 'Stage 2' mindreading in infancy is due entirely to the greater executive demands that are imposed by tasks of the latter sort. No new representational resources are involved.[16]

[15] Might the infants be representing not, *he thinks the ball is behind the screen*, but rather, *when he was present the ball was behind the screen*? This would require just noting the coincidence in time between the last presence of the agent and the location of the ball, not mindreading. But it is far from clear why such a representation should make the outcome (no ball behind the screen) in any way surprising or interesting, given that the infant itself saw the ball leave. In contrast, if infants have a general expectation that people's beliefs are likely to be true, or if they have some disposition to infer *P* from *she thinks that P*, then the absence of a ball will violate these expectations (while conforming to the infant's *own* expectations).

[16] Note that on some accounts of intentional content (specifically, inferential role theories; Block, 1986; Carey, 2009) the *content* of the representations of belief entertained by 7-month-olds may

5. Conclusion

I have argued that none of the dichotomies that have been used to characterize the significance of the recent infancy data are well supported. Infants are unlikely to be deploying a set of behavior-rules, nor representing others' thoughts in a non-propositional manner. And the present evidence suggests that mindreading in infancy doesn't undergo any deep transformation between Stage 1 and Stage 2 kinds of performance. On the contrary, I suggest that the best account of the existing data is that the domain-specific component of the mindreading system is available by around the middle of the first year of life. What changes over development are the interactions between this system and executive systems (together, no doubt, with elaboration of the information contained in the mindreading system resulting from the infant's own learning, including the acquisition of explicit concepts of truth and falsity). No new mechanisms are built or come online. And no deep changes in the representational resources available for mindreading take place thereafter.

Department of Philosophy
University of Maryland

References

Apperly, I. 2011: *Mindreading*. New York: Psychology Press.

Apperly, I. and Butterfill, S. 2009: Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116, 953–70.

Astington, J. and Jenkins, J. 1999: A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, 35, 1311–20.

Baars, B. 1988: *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Back, E. and Apperly, I. 2010: Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, 115, 54–70.

Baddeley, A. 2006: *Working Memory, Thought, and Action*. Oxford: Oxford University Press.

Baillargeon, R., He, Z., Setoh, P., Scott, R. and Yang, D. (in press). The development of false-belief understanding and why it matters. In M. Banaji and S. Gelman (eds), *The Development of Social Cognition*. Mahwah, NJ: Erlbaum.

undergo change and enrichment once those representations become appropriately connected with frontal planning systems and begin to be used to generate expectations of behavior. But this will be change of an incremental sort involving the very same content-bearing representations, not radical conceptual change in the sense of Carey (2009) or endorsed by defenders of the Stage 1 / Stage 2 distinction.

Baron-Cohen, S. 1995: *Mindblindness*. Cambridge, MA: MIT Press.

Barrett, L., Tugade, M. and Engle, R. 2004: Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, 130, 553–73.

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N. and Spelke, E. 2006: Non-symbolic arithmetic in adults and young children. *Cognition*, 98(3), 199–222.

Birch, S. and Bloom, P. 2004: Understanding children's and adult's limitations in mental state reasoning. *Trends in Cognitive Sciences*, 8, 255–60.

Block, N. 1986: Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, 10, 615–78.

Buttelmann, D., Carpenter, M. and Tomasello, M. 2009: Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112, 337–42.

Byrne, R. and Whiten, A. (eds) 1988: *Machiavellian Intelligence*. Oxford: Oxford University Press.

Byrne, R. and Whiten, A. (eds) 1997: *Machiavellian Intelligence II*. Cambridge: Cambridge University Press.

Call, J. and Tomasello, M. 1999: A nonverbal false belief task: the performance of children and great apes. *Child Development*, 70, 381–95.

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

Carlson, S., Moses, L. and 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.

Carlson, S., Moses, L. and Claxton, L. 2004: Individual differences in executive functioning and theory of mind: an investigation of inhibitory control and planning ability. *Journal of Experimental child Psychology*, 87, 299–319.

Carruthers, P. 2006: *The Architecture of the Mind*. Oxford: Oxford University Press.

Carruthers, P. 2009: An architecture for dual reasoning. In J. Evans and K. Frankish (eds), *In Two Minds*. Oxford: Oxford University Press.

Carruthers, P. 2011: *The Opacity of Mind*. Oxford: Oxford University Press.

Clements, W. and Perner, J. 1994: Implicit understanding of belief. *Cognitive Development*, 9, 377–95.

Csibra, G., Bíró, S., Koós, O. and Gergely, G. 2003: One-year-old infants use teleological representations of actions productively. *Cognitive Science*, 27, 111–33.

De Neys, W. 2006: Dual processing in reasoning: two systems but one reasoner. *Psychological Science*, 17, 428–33.

Dennett, D. 1987: *The Intentional Stance*. Cambridge, MA: MIT Press.

Dunbar, R. 1998: The social brain hypothesis. *Evolutionary Anthropology*, 6, 178–90.

Evans, J. 2008: Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–78.

Evans, J. and Over, D. 1996: *Rationality and Reasoning*. New York: Psychology Press.

Frankish, K. 2004: *Mind and Supermind*. Cambridge: Cambridge University Press.

Frankish, K. 2009: Systems and levels. In J. Evans and K. Frankish (eds), *In Two Minds*. Oxford: Oxford University Press.

Gallagher, S. 2001: The practice of mind: theory, simulation, or primary interaction? *Journal of Consciousness Studies*, 8 (5–7), 83–107.

Gallagher, S. 2004: Understanding interpersonal problems in autism: interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, and Psychology*, 11, 199–217.

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

Goldman, A. 2006: *Simulating Minds*. New York: Oxford University Press.

Gopnik, A. and Meltzoff, A. 1997: *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Heal, J. 1996: Simulation, theory, and content. In P. Carruthers and P.K. Smith (eds), *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Hrdy, S. 2009: *Mothers and Others*. Cambridge, MA: Harvard University Press.

Hutto, D. 2004: The limits of spectatorial folk psychology. *Mind & Language*, 19, 548–73.

Hutto, D. 2008: *Folk Psychological Narratives*. Cambridge, MA: MIT Press.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux.

Kahneman, D. and Frederick, S. 2002: Representativeness revisited: attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin and D. Kahneman (eds), *Heuristics and Biases*. Cambridge: Cambridge University Press.

Keenan, T. 1998: Memory span as a predictor of false belief understanding. *New Zealand Journal of Psychology*, 27, 36–43.

Kloo, D. and Perner, J. 2003: Training transfer between card sorting and false belief understanding: helping children apply conflicting descriptions. *Child Development*, 74, 1823–39.

Knudsen, B. and Liszkowski, U. 2012: 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy*, 17, 672–91.

Kosslyn, S. 1994: *Image and Brain*. Cambridge, MA: MIT Press.

Kovács, Á. 2009: Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, 12, 48–54.

Kovács, Á., Téglás, E. and Endress, A. 2010: The social sense: susceptibility to others' beliefs in human infants and adults. *Science*, 330, 1830–34.

Lakatos, I. 1970: The methodology of scientific research programs. In I. Lakatos and A. Musgrave (eds), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.

Leslie, A. 1994: ToMM, ToBy and Agency: Core architecture and domain specificity. In L. Hirchfeld and S. Gelman (eds), *Mapping the Mind*. Cambridge: Cambridge University Press.

Leslie, A. and Polizzi, P. 1998: Inhibitory processing in the false belief task: two conjectures. *Developmental Science*, 1, 247–53.

Leslie, A., Friedman, O. and German, T. 2004: Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences*, 8, 528–33.

Levelt, W. 1989: *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Lin, S., Keysar, B. and Epley, N. 2010: Reflexively mindblind: using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, 46, 551–6.

Low, J. 2010: Preschoolers' implicit and explicit false-belief understanding: relations with complex syntactical mastery. *Child Development*, 81, 597–615.

Low, J. and Watts, J. forthcoming: Attributing false-beliefs about object identity is a signature blindspot in humans' efficient mindreading system. *Psychological Science*.

McKinnon, M. and Moscovitch, M. 2007: Domain-general contributions to social reasoning: theory of mind and deontic reasoning re-explored. *Cognition*, 102, 179–218.

Milligan, K., Astington, J. and Dack, L. 2007: Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, 78, 622–46.

Neumann, A., Sodian, B. and Thoermer, C. 2009: Belief-based action anticipation in 18-month-old infants. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Denver, Colorado; April 2009.

Nichols, S. and Stich, S. 2003: *Mindreading*. Oxford: Oxford University Press.

Onishi, K. and Baillargeon, R. 2005: Do 15-month-olds understand false beliefs? *Science*, 308, 255–8.

Perner, J. 2010: Who took the cog out of cognitive science? Mentalism in an era of anti-cognitivism. In P. Frensch and R. Schwarzer (eds), *Cognition and Neuropsychology: International Perspectives on Psychological Science: Volume 1*. New York: Psychology Press.

Perner, J. and Ruffman, T. 2005: Infants' insight into the mind: how deep? *Science*, 308, 214–6.

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

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

Pylyshyn, Z. 2003: *Seeing and Visualizing*. Cambridge, MA: MIT Press.

Richerson, P. and Boyd, R. 2005: *Not By Genes Alone*. Chicago, IL: University of Chicago Press.

Rosenthal, D. 1986: Two concepts of consciousness. *Philosophical Studies*, 49, 329–59.

Ruffman, T., Garnham, W., Import, A. and Connolly, D. 2001: Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology*, 80, 201–24.

Sabbagh, M., Bowman, L., Evraire, L. and Ito, J. 2009: Neurodevelopmental correlates of theory of mind in preschool children. *Child Development*, 80, 1147–62.

Saxe, R. 2009: Theory of mind (neural basis). In W. Banks (ed), *Encyclopedia of Consciousness*. Cambridge, MA: MIT Press.

Saxe, R., Whitfield-Gabrieli, S., Pelphrey, K. and Sholz, J. 2009: Brain regions for perceiving and reasoning about other people in school-aged children. *Child Development*, 80, 1197–1209.

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

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

Slade, L. and Ruffman, T. 2005: How language does (and does not) relate to theory of mind: a longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology*, 23, 117–141.

Sloman, S. 1996: The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22.

Sloman, S. 2002: Two systems of reasoning. In T. Gilovich, D. Griffin and D. Kahneman (eds), *Heuristics and Biases*. Cambridge: Cambridge University Press.

Smolensky, P. and Legendre, G. 2006: *The Harmonic Mind* (2 volumes). Cambridge, MA: MIT Press.

Song, H. and Baillargeon, R. 2008: Infants' reasoning about others' false perceptions. *Developmental Psychology*, 44, 1789–95.

Song, H., Onishi, K., Baillargeon, R. and Fisher, C. 2008: Can an actor's false belief be corrected by an appropriate communication? Psychological reasoning in 18.5-month-old infants. *Cognition*, 109, 295–315.

Southgate, V., Chevallier, C. and Csibra, G. 2010: Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, 13, 907–12.

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

Sperber, D. and Wilson, D. 2002: Pragmatics, modularity, and mindreading. *Mind & Language*, 17, 3–23.

Stanovich, K. 1999: *Who is Rational?* Mahwah, NJ: Erlbaum.

Stanovich, K. 2009: *What Intelligence Tests Miss: The Psychology of Rational Thought*. New Haven, CT: Yale University Press.

Stanovich, K. and West, R. 2000: Individual differences in reasoning: implications for the rationality debate. *Behavioral and Brain Sciences*, 23, 645–726.

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

Träuble, B., Marinovic, V. and Pauen, S. 2010: Early theory of mind competencies: do infants understand others' beliefs? *Infancy*, 15, 434–44.

Tye, M. 1995: *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

Wellman, H. 1990: *The Child's Theory of Mind*. Cambridge, MA: MIT Press.

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

Wilson, T., Lindsey, S. and Schooler, T. 2000: A model of dual attitudes. *Psychological Review*, 107, 101–26.

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