

8-1-2010

Mindreading, Language and Simulation

Ryan C. DeChant
Georgia State University

Follow this and additional works at: http://scholarworks.gsu.edu/philosophy_theses

Recommended Citation

DeChant, Ryan C., "Mindreading, Language and Simulation." Thesis, Georgia State University, 2010.
http://scholarworks.gsu.edu/philosophy_theses/74

This Thesis is brought to you for free and open access by the Department of Philosophy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

MINDREADING, LANGUAGE AND SIMULATION

by

RYAN C. DECHANT

Under the Direction of Daniel Weiskopf

ABSTRACT

Mindreading is the capacity to attribute psychological states to others and to use those attributions to explain, predict, and understand others' behaviors. In the past thirty years, mindreading has become the topic of substantial interdisciplinary research and theorizing, with philosophers, psychologists and, more recently, neuroscientists, all contributing to the debate about the nature of the neuropsychological mechanisms that constitute the capacity for mindreading. In this thesis I push this debate forward by using recent results from developmental psychology as the basis for critiques of two prominent views of mindreading. First, I argue that the developmental studies provide evidence of infant mindreading and therefore expose a flaw in José Bermúdez's view that certain forms of mindreading require language possession. Second, I argue that the evidence of infant mindreading can also be used to undermine Alvin Goldman's version of Simulation Theory.

INDEX WORDS: Mindreading, Theory of mind, Social cognition, Propositional attitudes, False-belief tasks, Non-verbal false-belief tasks, Violation-of-expectation paradigm, Concept possession, Non-linguistic cognition, Second-order cognitive dynamics, Access consciousness, Metarepresentation, Simulation theory, Executive function, Inhibitory control

MINDREADING, LANGUAGE AND SIMULATION

by

RYAN C. DECHANT

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2010

Copyright by
Ryan Christopher DeChant
2010

MINDREADING, LANGUAGE AND SIMULATION

by

RYAN C. DECHANT

Committee Chair: Daniel Weiskopf

Committee: George Graham
Eddy Nahmias
Andrea Scarantino

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
August 2010

To Henry Loring Masters:

I wish you knew how much we all miss you.

ACKNOWLEDGEMENTS

A number of others contributed to this thesis, either directly or indirectly, and I'd like to recognize and thank them.

First, much thanks to Dan Weiskopf, who graciously and unhesitatingly agreed to advise the project, despite never having me in class, and despite the fact that he was supposed to be freed from all advising responsibilities this year. Always willing to meet and discuss things, Dan has a knack for striking a balance between encouragement and criticism that I find very instructive. I feel very fortunate to have had him as an advisor.

Thanks also to my thesis committee, George Graham, Eddy Nahmias and Andrea Scarantino. I've learned a lot from each of them during my time at Georgia State, and know I will continue to do so in the future.

A number of my fellow graduate students also deserve special acknowledgement. Paul Pfeilschifter was kind enough to let me crash (and party) at his house for a few nights during my prospective student visit to GSU, and ever since then I've admired him, both as a person and as a philosopher. Once I arrived on campus, Dan Burnston and Ben Sheredos lead by example. Trip Glazer, Reuben Stern and Brad Summers were always first-rate company, no matter whether we were talking shop, drinking beer or both. The same goes for Matt Duncan, who—along with his wife Megan—generously invited the whole crew over for dinner nearly every Thursday. Dylan Murray and Rush Stewart: you guys are reasonable fuckin' people, and damn good philosophers; I had a great year at 192 Mayson—despite the raccoons and foreclosure.

Lastly, I owe the most to my family. Mom, Dad, Bri: without you I wouldn't be where I am, so much thanks and love.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
INTRODUCTION	1
CHAPTER 1: MINDREADING, LANGUAGE, AND CONCEPT POSSESSION	4
1.1 THE POSSIBILITY OF NON-LINGUISTIC MINDREADING	7
1.2 THE ACTUALITY OF NON-LINGUISTIC MINDREADING	18
1.3 NON-LINGUISTIC MINDREADING AND CONCEPT POSSESSION	26
1.4 CONCLUSION	33
CHAPTER 2: MINDREADING, SIMULATION AND INHIBITORY CONTROL	35
2.1 GOLDMAN'S SIMULATION THEORY	38
2.2 SIMULATION THEORY AND INHIBITORY CONTROL	46
2.3 MINDREADING WITHOUT INHIBITORY CONTROL	55
2.4 CONCLUSION	66
BIBLIOGRAPHY	68

INTRODUCTION

“Mindreading,” or “theory of mind,” refers to the capacity to attribute psychological states to others and to use those attributions to explain, predict, and understand others’ behavior. Contemporary research on mindreading accelerated when Premack and Woodruff (1978) asked, “Does the chimpanzee have a theory of mind?” Since then mindreading has become the topic of substantial interdisciplinary experimentation and theorizing, with philosophers, psychologists and, more recently, neuroscientists all contributing to the debate about the nature of the neuropsychological mechanisms of mindreading. In this thesis I hope to move this debate forward by using recent results from developmental psychology as the basis for critiques of two prominent views of mindreading, and thereby limit the space of theoretical possibilities that we need to explore in the search for the correct account of mindreading.¹

In Ch. 1, I challenge José Bermúdez’s claim that it is impossible for non-linguistic organisms to engage in propositional attitude mindreading (i.e., to attribute propositional attitudes, like the belief that *p*, the desire that *q*, etc., and to use those attributions to explain, predict, and understand their behavior). I begin by arguing that Bermúdez’s attempt to establish a necessary link between propositional attitude mindreading and natural language possession fails because it equivocates between two notions of conscious accessibility. Then I support my argument by presenting results from a recent study by Scott and Baillargeon (2009), results that, I argue, overcome the “behavior-reading” objection raised by Povinelli in debates about primate mindreading and thus indicate that prelinguistic infants have the capacity for propositional

¹For the remainder of the essay, I will use “mindreading” instead of “theory of mind,” both for the sake of concision, and because it is more theoretically neutral with regards to the debate between “Theory Theorists” and “Simulation Theorists” that I will consider in Chapter 2.

attitude mindreading. Lastly, I consider the objection that we ought to describe the infants' capacity for propositional attitude mindreading in terms of nonconceptual content, since the infants cannot yet possess the concepts of the relevant propositional attitudes (e.g., the concept [belief]) because they are prelinguistic and concept possession requires language possession. I argue that the objection relies upon a misplaced distinction between conceptual and non-conceptual representation. The conclusion, then, will be that prelinguistic infants engage in substantive propositional attitude mindreading and that we therefore have no reason to think that non-linguistic organisms ipso facto cannot do so as well.

In Ch. 2, I turn to consider how the recent evidence of infant propositional attitude mindreading bears on the debate between “Simulation Theorists” and “Theory Theorists.” I argue that the developmental evidence reveals that simulative processes, as defined by Alvin Goldman (2006), the leading defender of Simulation Theory, cannot be responsible for what are agreed by both sides of the debate to be some of the central phenomena to be explained by any satisfactory theory of mindreading. On Goldman's version of Simulation Theory there can be no accurate propositional attitude mindreading in the absence of what cognitive scientists call “inhibitory control.” Moreover, Goldman rightfully explains the transition in children's performance on standard false-belief tasks between ages three and four in terms of the developmental emergence of inhibitory control. So, he must predict that children younger than approximately age three should be incapable of accurate propositional attitude mindreading. Until recently this prediction has been borne out. But, recent studies demonstrating that infants can pass non-verbal false-belief tasks—including the study by Scott and Baillargeon—indicate that children much younger than three years old have and employ the capacity for accurate

propositional attitude mindreading. So, I argue, given the strength of the evidence both for infant mindreading and for the claim that inhibitory control accounts for the transition in standard false-belief task, the most plausible conclusion to draw is that Goldman's Simulation Theory is false.

CHAPTER 1: MINDREADING, LANGUAGE, AND CONCEPT POSSESSION

In the thirty years since Premack and Woodruff (1978) asked the evocative question “Does the chimpanzee have a theory of mind?” an extensive amount of interdisciplinary research has aimed at determining what it is to have a basic understanding of the minds of others. Despite the substantial research, however, unequivocal evidence of so-called “mindreading” (or “theory of mind”) outside of the human species has been less than forthcoming. For instance, in a recent review of research on chimpanzee mindreading, Call and Tomasello (2008) explain that though psychologists have found evidence that chimps exhibit at least a rudimentary grasp of goal-directed behavior, they have yet to find any reason to think that chimps have an understanding of the underlying psychological states that give rise to that behavior. More skeptically, Penn and Povinelli (2007) conclude their review of such research with the claim that there is a complete “lack of evidence that non-human animals possess anything remotely resembling” a theory of mind.

Recently, José Bermúdez (2009; 2003) has offered an indirect explanation of this dearth of evidence. According to Bermúdez, such evidence in principle cannot be found because there are psychological prerequisites for mindreading that can only be met by organisms that possess a natural language. Thus, on his view, it is psychologically impossible for non-linguistic organisms to engage in mindreading, and empirical research aimed at testing whether they do so is misguided.

By contrast, my aim in this chapter is to defend the possibility of non-linguistic mindreading. I will begin by presenting and criticizing Bermúdez’s main argument against that possibility in §1.1. There I will argue that his argument relies upon a controversial premise

regarding the relationship between mindreading and consciousness, and that the sub-argument that he adduces in support of that premise equivocates between two notions of conscious accessibility, rendering it either invalid or unsound. Then, in §1.2 I will present the results from a recent study by Scott and Baillargeon (2009), and argue that they overcome the “behavior-reading” objection raised by Povinelli and others in debates over primate mindreading. If my argument is correct, the results indicate that prelinguistic infants have the capacity to engage in mindreading and therefore corroborate the claim that Bermúdez’s controversial premise is in fact false. Lastly, in §1.3 I will consider the objection that the developmental evidence of prelinguistic mindreading ought to be explained in terms of non-conceptual content, because the prelinguistic infants involved cannot yet possess the concepts of the relevant psychological states (e.g., the concept [belief]) since concept possession requires language possession. In response, I will argue that the objection is motivated by a misguided account of the distinction between conceptual and non-conceptual representation. The conclusion, then, will be that prelinguistic infants engage in quite sophisticated forms of mindreading, and that we therefore have no reason to think that non-linguistic organisms ipso facto cannot do so as well.²

Before proceeding, though, it will be important to rehearse two distinctions that Bermúdez introduces in order to precisify his claims. First, there is the distinction between “minimal” mindreading and “substantive” mindreading. According to Bermúdez, an organism engages in minimal mindreading when its behavior *systematically covaries* with the psychological states of the other organisms with which it interacts. In contrast, an organism

²Bermúdez is not the only theorist to have argued that mindreading is dependent upon natural language—see, for instance, de Villiers (2000), Garfield *et al.* (2001), and Smith (1996), among many others. Due to space constraints, though, here I will have to focus on his account alone.

engages in substantive mindreading when its behavior is *systematically causally dependent* on its *representations* of the psychological states of the other organisms with which it interacts.

Because his argument focuses on the psychological capacities required for forming such “metarepresentations,” Bermúdez’s conclusion is meant to apply only to substantive mindreading.³

Second, Bermúdez differentiates two kinds of substantive mindreading. Because substantive mindreading involves a causal dependency between an organism’s psychological states and a mindreader’s representations of those states, it requires the mindreader to “track” how those states interact with the other organism’s background psychological profile. And such tracking requires the mindreader’s being sensitive to the different causal roles of different types of psychological states. Thus, we should expect that *perceptual* mindreading (i.e., forming representations of another’s perceptual states or perspective) and *propositional attitude* mindreading (i.e., forming representations of another’s beliefs, desires, etc.) will involve different capacities. In particular, Bermúdez argues, we should expect that perceptual mindreading demands less psychologically of the mindreader than propositional attitude (PA) mindreading, because the causal relations between perceptual states and action is relatively straightforward when contrasted with the causal relations between PA’s and action. Given that the capacity to keep track of the complex causal roles of PA’s is, according to Bermúdez, the key feature of mindreading that requires language, his attempted repudiation of the possibility of non-linguistic mindreading is aimed only at PA mindreading.

³The term “metarepresentation” is frequently used to refer to any higher-order psychological state, including both the states involved in mindreading (thinking about others’ thoughts) and the states involved in “metacognition” (thinking about one’s own thoughts). However, in this paper I will not be concerned with metacognition, instead focusing exclusively on mindreading.

1.1 THE POSSIBILITY OF NON-LINGUISTIC MINDREADING

With these distinctions in mind, I can now evaluate Bermúdez's (2009) argument for the impossibility of non-linguistic PA mindreading (hereafter: "IMP"), which can be recapitulated as follows:

(P1.1) The representations involved in PA mindreading must be consciously accessible constituents of the mindreader's psychological life.

(P1.2) The representations involved in PA mindreading must exhibit the canonical structure of the propositions represented, and the canonical structure of a proposition is only exhibited when the proposition is represented in a linguistic format.

(C1.1) Therefore, the representations involved in PA mindreading must be natural language sentences, since natural language is the only consciously accessible form of linguistic representation.

(C1.2) Therefore, non-linguistic organisms ipso facto are unable to engage in PA mindreading.

Although one may find IMP *prima facie* persuasive, I think closer analysis demonstrates that in defending it Bermúdez faces a dilemma. Depending upon how he explicates the notion of "conscious accessibility" in P1.1 and C1.1, either the representations involved in PA mindreading do not have to be consciously accessible (which would make P1.1 false and IMP unsound), or they do have to be consciously accessible but only in a way that is independent of natural language (which would make the inference from P1.1 and P1.2 to C1.1 invalid). Or, at least that is what I will argue presently.

First, though, I want to grant P1.2. The psychological processes responsible for PA mindreading must be sensitive to the inferential relations between the PA's that are represented.

As Bermúdez (2009) explains, PA mindreading “requires working out the different possible inferential connections between those propositional attitudes and the agent’s background psychological profile, as well as the information that they [sic] have through perception about the distal environment” (159). And this requires that the mindreader represent others’ PA’s in a “format” that is sensitive to those relations. In Bermúdez’s terms, the representational format of meta-PA’s—that is, PA’s that represent or have as part of their content other’s PA’s—must exemplify the “canonical structure” of the propositions that are represented. And, at least for present purposes, I am willing to concede that only a linguaform system of representation can exhibit such sensitivity.⁴

Now what about P1.1? The support Bermúdez adduces for it is expressed in the following:

When a propositional attitude mindreader forms beliefs about the mental states of another agent, those beliefs are integrated with the rest of the mindreader’s propositional attitudes—with their beliefs about the distal environment, with their short-term and long-term goals, for example. This integration is required for the results of propositional attitude mindreading to feature in a creature’s practical decision-making. And, since practical decision-making takes place at the conscious level, beliefs about the mental states of other agents must also be consciously accessible.⁵

Here Bermúdez appears to be trying to establish a necessary link between forming meta-PA’s and exercising a certain type of decision-making, and, in turn, a necessary link between that type of decision-making and conscious accessibility. As I see it, the argument, which I will call “NCA” for “necessarily consciously accessible,” can be reconstructed as follows:

(P2.1) If a psychological process takes place at the level of consciousness, then the representations that are capable of playing a role in that process must be consciously

⁴For relevant discussion see Camp (2007) and Rescorla (2009).

⁵Bermúdez (2009), p. 159; see also pp. 150-151 for roughly the same argument.

accessible.

(P2.2) Practical decision-making takes place at the level of consciousness.

(C2.1) Therefore, the representations that are capable of playing a role in practical decision-making must be consciously accessible.

(P2.3) If meta-PA's are capable of playing a role in the mindreader's practical decision-making, then they must undergo cognitive integration with a mindreader's other PA's.

(P2.4) Meta-PA's must undergo cognitive integration with a mindreader's other PA's.

(C2.2) Therefore, meta-PA's must be consciously accessible.⁶

Now before assessing the plausibility of NCA's premises, it is important to recognize that as it stands the argument is invalid, since it ignores the possibility of meta-PA's that are not capable of playing a role in practical decision-making. The problem is that P2.3 is the claim that cognitive integration is a *necessary* condition for a meta-PA's being capable of playing role in practical decision-making. Yet, if a meta-PA's undergoing cognitive integration is only necessary and not sufficient for it's being capable of playing role in practical decision-making, then it is only the meta-PA's that are capable of playing a role in such decision-making—rather than meta-PA's *simpliciter*—that must be consciously accessible.

So, for NCA to be valid, P2.3 needs to be replaced with the claim that undergoing cognitive integration is *sufficient* for a representation's being able to play a role in practical decision-making:

(P2.3*) If meta-PA's undergo cognitive integration, then they must be capable of

⁶One will notice that I have reconstructed Bermúdez's sub-argument as making modal claims, and may be worried that the passage quoted above does not necessitate such a reconstruction. This worry should be assuaged by the fact that only if the sub-argument is understood in this way will it entail P1.1 of IMP. See pp. 150-151 of Bermúdez (2009) for further support of this interpretation.

playing a role in the mindreader's practical decision-making.

Otherwise, the claim that meta-PA's must be cognitively integrated would not entail anything about the potential role of meta-PA's in practical decision-making, and thus would not entail anything about their conscious accessibility. Here I am not trying to quibble over irrelevant details. My impression is that Bermúdez's vacillation on this point is the source of the problems that ultimately undermine his case against the possibility of non-linguistic mindreading. So, for what follows, I propose to trade being literal for being charitable, and replace P2.3 with P2.3*.

Now how secure are NCA's premises? To start, P2.1 seems to stand somewhere between innocuous and tautologous, so I am willing to grant it. And P2.4 seems like it must be accepted by anyone who is committed to representationalism about PA's in general.⁷ The reason is that it would be explanatorily superfluous to say that an organism's having a certain PA amounts to its standing in the appropriate relation to a certain mental representation yet deny that the representation must be able to interact with (at least some) other representational states in order to guide the organism's actions. As Bermúdez (1998) explains, "there must be [psychological] pathways enabling a given representational state to connect up with other states, both representational and motivational. There must be *cognitive integration* of the relevant states" (91; emphasis in original). Otherwise, it seems that PA's would be causally inefficacious. So, I will grant that meta-PA's must undergo cognitive integration.

P2.2 and P2.3*, however, are another story. As I see it, there are two possible ways to construe the notion of practical decision-making (hereafter: "PDM") at work here, the relevant

⁷Although this view of PA's is controversial, because Bermúdez is committed to it, I am going to assume it is true for the purposes of this essay. See Clark (2001), amongst many others, for critical discussion.

difference between the two being the cognitive sophistication required for engaging in each.⁸ The first, less cognitively sophisticated construal identifies PDM as the psychological process involved in choosing a particular action from a set of alternatives because of the comparative desirability of its consequences. As Bermúdez (2003) puts it, PDM just is “selection between different possible courses of action that is grounded on an assessment of the likely consequences that those different possible courses of action will have.” What, then, is required in terms of cognitive architecture for this type of PDM? According to Bermúdez, “to say that an action is being performed on consequence-sensitive grounds implies ... that the agent has made an assessment of those consequences, on the basis of a belief about the outcome that that action is likely to have... Decision-making only takes place, in other words, at a level where instrumental beliefs are available” (*ibid.*). So, if PDM just is acting on the basis of an evaluation of a set of possible actions, it requires (a) representing the contingencies that hold between particular actions and the state of affairs that are expected to be the outcomes of those actions, and, further, (b) acting on the basis of an evaluative comparison of those representations.

In contrast, Bermúdez’s (2009) use of the notion of PDM while defending IMP can be construed in a second, more cognitively-demanding way. On this view, PDM requires “explicitly” representing and reasoning about the entailment relations between various thoughts, possible courses of action, and the predicted outcomes of those actions. This suggests that Bermúdez intends PDM to be a form of what Clark (2009; 1998) has called “second-order

⁸No doubt there are, strictly speaking, innumerable possible ways of construing PDM, but it seems to me that the two construals that I offer are the most plausible given both the aim of Bermúdez’s argument and what he explicitly says.

cognitive dynamics.”⁹ According to Clark, second-order cognitive dynamics involves explicit, critical reflection on one’s own thought processes. Yet, as Bermúdez (2003) himself explains, if PDM is a form of second-order cognitive dynamics, the only creatures that can engage in PDM are those that have the capacity for “explicit consideration of the formal logical relations between thoughts” on the basis of an understanding of the basic principles of inductive and deductive reasoning—that is, those that can engage in what he calls “reflective doxastic modification” (170). So, on this second construal, having the capacity for PDM requires much more than the ability to form instrumental beliefs and act on the basis of those beliefs. Instead, it requires a certain form of explicit epistemic self-management and is therefore much more psychologically demanding.

Which of these two construals of PDM does Bermúdez intend to be operative in NCA? I am not entirely sure, for in different places he says things suggesting that he favors each. However, I see no need to engage in sustained exegesis, because I am going to argue that in either case he will be unable to establish the impossibility of non-linguistic PA mindreading. On the one hand, if Bermúdez construes PDM merely as instrumental reasoning, NCA will indeed turn out to be sound and meta-PA’s will have to be consciously accessible, but only in a way that renders his main argument (IMP) for the impossibility of non-linguistic mindreading invalid. On the other hand, if he construes PDM as a form of second-order cognitive dynamics, NCA will be unsound and therefore provide no support for the controversial premise of IMP, P1.1.

To see why, consider the first horn of the dilemma. If PDM is instrumental reasoning, then it does seem plausible that meta-PA’s must at least be *capable* of playing a role in it. The

⁹See Bermúdez (2009, pp. 148-152; 158-159). Also, see Bermúdez (2003, Ch. 9-10), where explicitly aligns his notion of PDM with Clark’s notion of second-order cognitive dynamics.

reason is that by current assumption all PA's must undergo cognitive integration and it is arguable that undergoing cognitive integration is sufficient for being capable of playing a role in instrumental reasoning. Suppose that an organism *O* forms the belief that *p*. It follows that *O*'s representation of *p* must be able to causally interact with the representations that are the constituents of (at least some of) *O*'s other PA's. If that is the case, though, there is no reason to think that *O*'s representation of *p* would be unable to play a role in *O*'s instrumental reasoning *if O is able to engage in such reasoning*. And P2.3* only claims that a PA's undergoing cognitive integration is sufficient for it to *possibly* figure in PDM, and does not claim that it *actually* has to do so. Thus, if PDM is instrumental reasoning, meta-PA's are linked to it in the way required by P2.3*.

Adopting the weaker construal of PDM comes with a burden though. Recall that I've granted that if a psychological process takes places at the level of consciousness, the representations that are capable of playing a role in that process must be consciously accessible (P2.1). Now suppose we grant that PDM takes places at the level of consciousness (P2.2). It follows that the representations that are capable of playing a role in the PDM of non-linguistic organisms are consciously accessible. But if the representations of non-linguistic organisms can be consciously accessible, then the inference in IMP from P1.1 and P1.2 to C1.1 is fallacious, because it begs the question against a defender of the language of thought hypothesis. According to the language of thought hypothesis, thinking *just is* the causal interactions among an internal system of linguaform representations. The system is considered to be language-like because the inferential relations among the contents of the representations are supposed to be reflected by their formal or "syntactic" properties (or, in Bermúdez's terms, because the representations'

formal properties reflect their “canonical structure”), and because the processes that govern the causal interactions among the individual representations are supposed to be sensitive to those formal properties. In this way, the causal interaction among the representations can be responsive to the inferential relations among the contents of those representations.¹⁰ And, although it is a truism that non-linguistic organisms do not possess a natural language, it very much remains an open empirical question whether those organisms’ cognitive capacities are underwritten by such a system of linguaform representations. So, Bermúdez cannot assume that they do not. But, without that assumption C1.1 doesn’t follow from P1.1 and P1.2, if Bermúdez maintains that the representations that can play a role in PDM are consciously accessible. Thus, if PDM just is instrumental reasoning, IMP is invalid and does not provide any reason to think that non-linguistic organisms are ipso facto incapable of PA mindreading.

Bermúdez must therefore opt for the stronger construal of PDM if IMP is going to stand as a serious challenge to the possibility of non-linguistic PA mindreading. But this just raises the second horn of the dilemma: if PDM is a form of second-order cognitive dynamics, then there is no reason to assume that all meta-PA’s must be capable of playing a role in PDM. Recall the cognitive sophistication required for second-order cognitive dynamics. If a PA is to play a role in second-order cognitive dynamics, then it must be available for “explicit” evaluation on the basis of its logical and inferential relations with other thoughts, evidence, and behavior. Suppose we grant Bermúdez that such “reflective doxastic modification” cannot occur without the representations involved being natural language sentences, expressed either in overt or inner speech. It would follow that the representations that are capable of playing a role in PDM must

¹⁰See Fodor (1975) and Camp (2009b) for more on the language of thought hypothesis.

be natural language sentences. But, then there would be no reason at all to think that a PA *must* be able to play a role in PDM. This becomes clear when one considers all the organisms that certainly do not have the capacity for reflective doxastic modification—which, by current assumption, includes all non-linguistic organisms—but nevertheless exhibit sophisticated forms of behavior that can only be understood in terms of intentional explanations (viz., explanations that appeal to the organism's PA's).

To take what might be the most extreme example, there is strong empirical evidence that many of the more complex behaviors of insects cannot be accounted for without attributing PA's to them. For instance, Carruthers (2009; 2006) and Tetzlaff and Rey (2009) argue that the navigational skills of honeybees require ascribing to the bees a complex system of representations that have as their content numerous properties of the bees' environment, including the location of their nest, the distance between their nest and multiple feeding sites and landmarks, and the relative position of the sun. As Carruthers (2009) encourages, “the data warrant the claim that bees possess both belief-like states and desire-like states that interact with one another in simple practical inferences to select and guide behavior; and that the belief-like states possess a component structure, containing symbols that refer to various landmarks and substances as well as encoding the distances and directions between them” (93).

Less controversially, we now have very good reason to think that prelinguistic hominids were genuinely cognitive creatures. The best evidence suggests that such hominids engaged in sophisticated forms of tool construction, used their tools in long-distance, cooperation-based hunting expeditions that were planned in accordance with the seasons, and very likely exhibited the social coordination required to migrate from Africa to myriad sites throughout Eurasia—all

prior to the evolution of language. And there is near unanimous agreement that such behaviors cannot be satisfactorily explained if one refrains from attributing to the hominids the capacity for both representing the world and for employing at least rudimentary forms of instrumental reasoning—that is, if one refrains from attributing to them PA's.¹¹

Moreover, such examples can be multiplied repeatedly. And in each case, even though the organisms are non-linguistic—and by current assumption thereby lack the capacity for reflective doxastic modification—their behavior can only be understood by attributing to them states with intentional content.¹² Thus, if Bermúdez opts for the stronger construal of PDM, cognitive integration is not sufficient for being able to play a role in PDM and, as a result, NCA is rendered unsound.

Stepping back, we can see exactly where things go wrong for Bermúdez. IMP assumes that an organism must have conscious access to the representations that play a role in its PA mindreading (P1.1). But, in order for IMP to be valid, the operative notion of conscious accessibility must involve a necessary link with being represented in natural language. And, according to Bermúdez (2009), an organism engages in substantive PA mindreading whenever its “behavior is caused by (and hence can be explained by appeal to) how it represents the mental states of others” (148). So, the claim that the representations that play a role in PA mindreading must be consciously accessible requires justification. Otherwise, the possibility would remain

¹¹See, for instance, the essays in Carruthers and Chamberlain (2000). And for philosophical discussion of further examples of the sophistication of non-linguistic thought, see Beckoff, Allen and Burghardt (2002), Camp (2009a), Carruthers (2006), and Lurz (2009).

¹²One may want to deny the assumption that the capacity for reflexive doxastic modification requires possessing a natural language. But doing so would not help Bermúdez, since he needs it to undermine the possibility of non-linguistic PA mindreading. Also, although one may object to the idea that the behavior non-linguistic organisms requires genuine intentional explanation, this objection is not available to Bermúdez; his (2003) book was a sustained attempt to defend the claim that numerous behaviors exhibited by non-linguistic organisms can only be understood in intentional terms. So, for present purposes, the objection can be set aside.

that non-linguistic organisms could engage in PA mindreading through the use of representations that were not consciously accessible in the operative sense.

Now Bermúdez is aware that P1.1 stands in need of justification, and he puts forth NCA to provide that justification. However, NCA relies on the notion of PDM, which is ambiguous between two construals. If Bermúdez opts for the weaker construal—and PDM just is instrumental reasoning—then NCA turns out to be sound, but the only type of conscious accessibility that it underwrites involves what Block (2007) calls “access consciousness.” According to Block, a representation is access conscious if it can “be used freely as a premise in reasoning, according to the capabilities of the reasoner [and] to be used freely for control of action” (144). Thus, a representation’s being access conscious is not sufficient for its “vehicle” to be a natural language sentence, since the representations that play a role in rationally guiding the behavior of non-linguistic organisms are access conscious.

For that reason, Bermúdez must opt for the stronger construal of PDM in order for NCA to corroborate P1.1 in a way that retains the validity of IMP. Doing so, however, creates a psychological divide between cognitive integration and PDM, since on the stronger construal PDM requires reflective doxastic modification, and since many organisms have PA’s that undergo cognitive integration—and thus have PA’s that are capable of playing a role in PDM—despite lacking the capacity for reflective doxastic modification. And if such a divide exists, P2.3* is false and NCA is rendered unsound and insufficient for justifying P1.1. Therefore, no matter what Bermúdez intends PDM to be, his argument does not establish the claim that non-linguistic organisms are ipso facto incapable of PA-mindreading.

1.2 THE ACTUALITY OF NON-LINGUISTIC MINDREADING

To support this conclusion, I now want to consider a recent developmental study by Scott and Baillargeon (2009), and argue that the results can only be explained by attributing to the infant subjects the capacity for substantive PA mindreading. If my argument is correct, the results further corroborate my claim that Bermúdez’s premise (P2.3*) that if meta-PA’s undergo cognitive integration, then they must be capable of playing a role in the mindreader’s practical decision-making is false. The reason being that explaining the infants’ behavior in terms of PA-mindreading entails attributing to them meta-PA’s that undergo cognitive integration, even though the infants lack the capacity for reflective doxastic modification and thus lack the capacity for PDM. The results therefore provide additional reason to think that Bermúdez’s controversial premise (P1.1) that the meta-PA’s involved in PA mindreading must be consciously accessible is false, since the representations involved in the infants’ PA mindreading do not play a role in reflective doxastic modification and therefore are not consciously accessible in the way that the validity of IMP requires.¹³

However, given that the Scott and Baillargeon study is rather nuanced, I want to first briefly consider the experimental method standardly used to decide whether children have the capacity for PA mindreading—what is known as the “false-belief task.”¹⁴ The task typically involves an experimenter telling a child subject a story about two characters—often named “Sally” and “Anne”—while it is acted out by puppets. In the story, Sally and Anne are together

¹³Here I am presenting the argument in terms of the second, stronger construal of PDM. Alternatively, one could assume the first construal of PDM, according to which PDM just instrumental reasoning. In that case the infants’ meta-PA’s *would* be capable of playing a role in PDM and by NCA *would* be consciously accessible. This would show that P1.1 is true. But it would also make IMP invalid.

¹⁴See Wimmer and Perner (1983) and Gopnik (1993). Also, see Wellman et al. (2001) for a meta-analysis of studies using the standard false-belief task.

in a room when Sally places her toy in a drawer and leaves. Then, while Sally is gone, Anne moves the toy to another location. Having watched Anne move the toy, the subject is asked to say or point to where Sally will look for her toy when she returns.

Although the specific results vary with each experiment, the general finding has been that 4-year-olds typically pass the false-belief task, answering that Sally will look for her toy in the drawer where she left it and thus still believes it to be, whereas 3-year-olds typically fail the task, answering that Sally will look for the toy where it is actually located. And although the details of specific accounts vary, the traditional interpretation of these results has been that only 4-year-olds, and not 3-year-olds, have sufficiently developed mindreading capacities to understand that other agents' actions are caused by their beliefs (which can be either true or false) rather than the way things actually are.

Recently, though, the legitimacy of the standard false-belief task as an appropriate method for testing children's understanding of belief—and with it the traditional interpretation just mentioned—have come into question.¹⁵ The recent critiques raise issues regarding the performance-competence distinction, emphasizing the heavy linguistic, memory, and attentional demands of the task. And they have gained empirical support from studies in which, after being asked where Sally will search for her toy, children as young as 2 years old *look* at the correct location, despite *saying* the incorrect location.¹⁶ The concern, then, is that subjects younger than 4 have an understanding of belief that goes unnoticed in the standard false-belief tasks because it is masked by the task's undue performance demands. Thus, the capacity for PA mindreading may be present earlier than the results of such tasks suggest.

¹⁵See, for instance, Birch & Bloom (2004; 2007).

¹⁶See Clements & Perner (1994) and Garnham & Ruffman (2001).

It was with this concern in mind that Scott and Baillargeon (2009) developed a *non-verbal* false-belief task to use with 18-month-old prelinguistic infants.¹⁷ The non-verbal version of the task employed the “violation-of-expectation” paradigm, which is based upon the presupposition that infants direct more attention to events that violate their expectations, and thus should reliably look longer at scenes where something inconsistent with what they expect occurs.¹⁸ By systematically manipulating the features of scenes presented to the infants, experimenters the use the violation-of-expectation-paradigm are able to determine exactly the properties of the scenes to which the infants are cognitively sensitive. Thus, by appropriately manipulating the *psychological states* of the individual in the scenes with which their infants were presented, Scott and Baillargeon were able to use the paradigm to test whether infants’ expectations about how others will act are formed on the basis of an understanding of others’ psychological states.

The study began by presenting the infants with a set of four familiarization trials in which

¹⁷In fact, this study is only the most recent in a series by Baillargeon and colleagues, all of which use non-verbal versions of the task. Onishi and Baillargeon (2005) were the first to employ a non-verbal false-belief task and find evidence of prelinguistic PA mindreading, and others have since either replicated or extended their findings: Surian *et al.* (2007) found very similar results, except with 13-month-olds; so did Southgate *et al.* (2007), except they tested how long infants looked *in anticipation of*—as opposed to *in response to*—an agent’s action, suggesting that infants were not merely passively responding to that with which they were presented. Here I focus on the Scott and Baillargeon (2009) study because it shows that infants have more sophisticated PA mindreading capacities than the previous studies, and, as I will argue below, because it excludes the possibility of discounting the results as only warranting attributing to the infants the capacity for *behavior-reading* or *perceptual* mindreading, rather than *PA* mindreading.

¹⁸This paradigm has been used widely in developmental psychology (Baillargeon 2004), and more recently in comparative psychology (Hauser & Spaulding 2006), in particular for testing subjects’ knowledge of the principles that govern the behavior of physical objects. The guiding assumption is that if the subjects have an understanding of the principle being tested, they should look relatively longer when presented with an “impossible” (and hence unexpected) scene in which an object appears to violate the principle, and look for a relatively briefer period when presented with a “possible” (and hence expected) scene in which an object behaves in accordance with the principle. By contraposition, if the subjects show no difference in looking time between “impossible” and “possible” scenes, it is assumed that they have no understanding of the relevant principle and do not form the appropriate expectations. See Baillargeon (2000) for a defense of the method against criticisms, including the objection that subjects’ responses are due to contrasts in stimulus familiarity.

they watched an “agent” (i.e., an adult confederate) interact with two visually identical toy penguins, one of which could be disassembled into two pieces, the other of which could not. The first two trials proceeded as follows: first the infants saw the agent sitting at a table with two platforms on top of it; then they witnessed a gloved experimenter’s hand place the one-piece penguin on one of the platforms, and the two halves of the disassembled two-piece penguin on the other platform; lastly, they watched the agent hide a key in the bottom half of the two-piece penguin, put the top half and the bottom half together, and place the assembled penguin back on its platform. The second two familiarization trials were exactly the same, except that the locations of the penguins were reversed, and each penguin was placed in a shallow box rather than on a platform. This provided the infants with evidence that each penguin could appear in multiple positions and in multiple containers.

Next, the infants received one of two test conditions. In the false-belief condition, they watched while the agent was absent as the gloved hand assembled the two-piece penguin, placed it under a large transparent box, and then covered the one-piece penguin with a large opaque box; then the agent entered the room with her key in hand and reached for either the transparent or opaque box. In the true-belief condition, everything was the same, except that the agent was present the entire time. Across both conditions, half of the infants saw the agent reach for the transparent box, while the other half saw her reach for the opaque box.

This experimental design enabled Scott and Baillargeon to reason as follows. If the infants had the capacity for PA mindreading, then in the *false*-belief condition they should have recognized that the agent was not present while the gloved hand assembled the two-piece penguin and placed it under the transparent box; and they should have recognized that upon

entering the scene the agent would therefore mistake the assembled two-piece penguin under the transparent box as the one-piece penguin and come to believe falsely that the two-piece penguin was under the opaque box. So, they should have expected the agent to reach for the opaque box. In contrast, in the *true*-belief condition, the infants should have recognized that the agent *did* witness the gloved hand assemble the two-piece penguin and place it under the transparent box. So, they should have expected her to reach for the transparent box. Scott and Baillargeon thus predicted that if they were able to form the appropriate expectations, the infants in the false-belief condition ought to express surprise upon witnessing the agent reach for the transparent box, whereas the infants in the true-belief condition ought to express surprise upon witnessing the agent reach for the opaque box.

And, indeed, the results confirmed their prediction: the infants in the false-belief condition looked significantly longer when the agent reached for the transparent box, presumably violating their expectation that she would mistakenly reach for the opaque box, where she falsely believed the toy to be; and the infants in the true-belief condition looked significantly longer when the agent reached for the opaque box, presumably violating their expectation that she would correctly reach for the transparent box, where the toy was actually located.

Does this indicate that the infants have a capacity for substantive PA mindreading? By Bermúdez's standards, almost certainly. In the false-belief condition, in order to correctly expect the agent to reach for the opaque box, the infants had to (*i*) recall from the familiarization trials that the agent was aware that both the one-piece and two-piece penguins were present, and (*ii*)

that she wanted to hide her key in the two-piece penguin.¹⁹ They also had to (*iii*) recognize that the agent expected the two-piece penguin to be unassembled, and (*iv*) that this expectation would cause the agent to falsely believe that it was the one-piece penguin that was perceivable under the transparent cover. Then they had to (*v*) realize that the agent's false belief that the penguin under the transparent cover was the one-piece penguin would cause her to falsely infer that the two-piece penguin was under the opaque box. Lastly, they had to (*vi*) understand that the agent's belief about the location of the two-piece penguin and her desire to hide her key in that penguin would cause her to reach for the opaque box. Similarly, in the true-belief condition, the infants had to (*i*) recognize that the agent wanted to hide her key in the two-piece penguin, and (*ii*) that because she was present the entire time, she correctly believed that the two-piece penguin was located under the transparent box. Then they had to (*iii*) understand that her belief and desire would cause her to reach for the transparent box. In both conditions, if the infants were not responsive to any one of these intervening variables, they would not have been able to form the correct expectations. As Scott and Baillargeon (2009) put it, successfully completing the task required the infants to attribute to the agent "a complex, interlocking set of motivational states, reality-congruent informational states [i.e. true beliefs], and reality-incongruent informational states [i.e. false beliefs]" (1180).

¹⁹A follow-up control experiment adjudicated between the interpretation that the infants attributed to the agent the desire to hide her key in to the two-piece penguin and the alternative interpretation that the infants ignored the key and simply attributed to the agent a preference for the two-piece penguin. It had two conditions. The first was identical to the false-belief condition in the first experiment, and the second was exactly the same except that the agent performed her actions without a key. If the infants only attributed to the agent a preference for the two-piece penguin, then they should have behaved identically across the key and no-key conditions. But they didn't. The infants in the key condition replicated the results from the original false-belief condition, whereas the infants in the no-key condition exhibited no difference in looking time between seeing the agent reach for the opaque box and seeing the agent reach for the transparent box. This supports the view that in the original false-belief condition the infants attributed to the agent a desire for a particular state of affairs (viz., that her key be hidden in the two-piece penguin) rather than a preference for an object (viz., the two-piece penguin).

Now perhaps one may object that this interpretation of the results attributes to the infants an unnecessarily rich set of representational capacities. For instance, in response to studies that purport to establish the existence of mindreading in non-human primates, Povinelli and colleagues have argued that those studies' results do not require attributing to the primate subjects the capacity for mindreading. Instead, they can be explained in terms of the primates' engaging in sophisticated forms of *behavior-reading*.²⁰ The reason, they argue, is that all such studies only require the subjects to (a) observe others' overt behavior and physical relations to the environment and (b) represent either learned or innate rules for predicting future behavior on the basis of those observations (rather than on the basis of an understanding of others' psychological states). Therefore, Povinelli argues, it is explanatorily superfluous to ascribe to the subjects the capacity for mindreading. And perhaps one might think that a similar analysis could be given for the Scott and Baillargeon study. Indeed, although the infants' behavior exhibited a sort of systematic causal dependence upon the psychological states of the agent, perhaps it only did so in virtue of the fact that the infants represented the agent's *behavioral* tendencies and the fact that those tendencies themselves in turn exhibited a sort of systematic causal dependence upon the agents' psychological states. If so, the results would only warrant the claim that infants engage in what Bermúdez defined as *minimal* mindreading. And that would eliminate any threat to his claim that non-linguistic mindreading is impossible, which concerns only *substantive* mindreading.

However, Povinelli's behavior-reading analysis cannot account for Scott and Baillargeon's results. The reason is that the results cannot be explained solely in terms of the

²⁰See Povinelli and Vonk (2004) and Penn and Povinelli (2007).

infants' keeping track of the *observable* relations between the agent's behavior and the environment. The infants did not, for instance, behave as if they expected the agent to reach where the two-piece penguin was *actually* located. Indeed, what is so telling about the results is that in the false-belief condition the infants expected the agent to mistakenly reach for the opaque box, *away* from the actual location of the two-piece penguin. Similarly, the infants did not behave as if they expected the agent to reach towards the location to which she had *previously* visually oriented or reached, since in the familiarization trials the infants observed the agent reach both to her left and to her right. Nor did the infants behave as if they merely expected the agent to reach where she had *last* visually oriented or reached (which was to her left in the fourth familiarization trial); such an expectation would lead the infants in the true-belief condition to look longer when the agent reached to her right for the transparent box, and the exact opposite was found.

Instead, in order to successfully complete the task, the infants in the false-belief condition had to recognize that the agent would see the two-piece penguin *as the one-piece penguin* and then *infer* mistakenly that the two-piece penguin must therefore be in the opaque box. And in the true-belief condition they had to recognize that because the agent was present the whole time, she would see the two-piece penguin *as the two-piece penguin*, even though it *appeared* as if it was the one-piece penguin. In other words, successfully completing the task thus required the infants to recognize regularities between the agent's behavior and *how she represented* the environment. The results therefore indicate that the infants were engaging in substantive PA mindreading, and that (*contra* Bermúdez) their capacity to do so was not in anyway dependent

upon natural language.²¹

1.3 NON-LINGUISTIC MINDREADING AND CONCEPT POSSESSION

At this point, someone sympathetic to Bermúdez’s claim that substantive non-linguistic PA mindreading is impossible may be moved by something like the following objection:

“Although the results from the non-verbal false-belief tasks indicate that we ought to attribute to the infants intentional states that have as part of their content the psychological states of others, this does not mean that it is appropriate to ascribe to them *genuine* meta-PA’s. This is because the following two principles are true:

Possession: For an organism *O* to genuinely have a PA, *O* must possess the concepts that are the constituents of the content of that PA.

Priority: Concept possession is constitutively linked with possession of a natural language in such a way that non-linguistic creatures cannot possess concepts.

And it follows from Possession and Priority that no non-linguistic organism could ever have the capacity for genuine meta-PA’s because such organisms ipso facto cannot possess the requisite PA concepts (e.g., the concept [belief], the concept [desire], etc.). In other words, although it appears as if we will have to attribute to the infants *proto*-meta-PA’s in order to explain their behavior, it does not follow that we must attribute to them full-fledged conceptual meta-PA’s, since that would require that they possess a language. Instead, we ought to explain their behavior

²¹ One may be concerned by the fact that the infants in the Scott and Baillargeon study were in the process of acquiring their first language, and thus do not, strictly speaking, count as *non*-linguistic organisms. However, the infants clearly constitute non-linguistic organisms in the sense that is relevant to Bermúdez’s argument. Bermúdez thinks that PA mindreading requires expressing in overt or inner speech natural language sentences that express the relevant metarepresentative content, and 18-month-old infants are not even close to that stage in language development. Indeed, infants do not acquire propositional attitude verbs until around their third birthday (Papafragou *et al.* 2007; Wellman & Bartsch, 1995). Moreover, the infants in the study I discussed are by no means the youngest that have successfully completed non-verbal false-belief tasks; Surian *et al.* (2007) found evidence of false-belief comprehension in infants as young as 13 months.

by attributing to them states with *non-conceptual* content.”²²

Although I cannot consider this objection in the detail it ultimately deserves here, I want to offer some reasons to think it cannot be sustained.²³ To begin, one must recognize that as Bermúdez and others has developed it, the notion of non-conceptual content has primarily been used to explain certain unique features of the content of perceptual experience, features that are supposed to contrast it with the content of propositional attitudes. As Bermúdez (2007) himself explains, “nonconceptualism is committed both to the thesis that perception is concept-independent and to the claim that this concept-independence is a function of the *distinctive type of content enjoyed by perceptual experiences*” (69; emphasis added).²⁴ On this view, “perceptual experiences can have richness, texture, and fineness of grain that beliefs do not and cannot have” (1998, 51). Thus, the existence of non-conceptual content must be acknowledged because the character of perceptual experience cannot be captured solely in terms of the more discrete, structured content that comes with conceptual representation.

So, if we assume this notion of non-conceptual content, explaining the mindreading

²²It is evident that Bermúdez would be likely to pursue this line of response; indeed, he explicitly endorses and defends both Possession and Priority in Bermúdez (1998).

²³It should be noted that the objection raises the bar for being considered a “genuine” PA-mindreader beyond what Bermúdez (2009) set as the threshold for substantive mindreading. Recall that “a creature engages in substantive mindreading when its behavior is systematically dependent upon its representations of the psychological states of other participants in the interaction” (148). Thus, the objection is, in effect, a concession that substantive PA mindreading is possible in the absence of language.

²⁴Bermúdez (2007) is quite explicit about the inherent link between non-conceptual content and the distinctive features of perceptual experience: on his view, “the core of the idea of nonconceptual content is a modal claim to the effect that perceivers *need not* possess concepts corresponding to everything that they are capable of perceptually discriminating” (69); similarly, he claims that “the central impetus for legitimating a notion of nonconceptual content has come from the study of perceptual experience... [T]heorists have been attracted to nonconceptual content by the thought that the richness and grain of perceptual experience is not constrained by the concepts that a perceiver might or might not possess” (1998, p. 50). Bermúdez (2007) is adamant about this connection because he denies the plausibility of trying to capture the distinction between conceptual and non-conceptual representations by saying conceptual and non-conceptual states have the *same type of content*, but are different because of the way that each state *functions* in an organism’s psychology—what Heck (2007) has called the “state” view of non-conceptual content.

capacities of infants in terms of non-conceptual content would require explaining their behavior solely on the basis of how they perceive things, refraining entirely from attributing to them any PA's. And when considering results from previous studies that employed the violation-of-expectation paradigm, that is exactly the strategy that Bermúdez (1998) has adopted. For instance, numerous studies have used the paradigm to test infants' understanding of the basic principles that govern the behavior of physical objects.²⁵ And Spelke (1988), along with others, has argued that the results of such studies warrant the conclusion that the infants possess the concept [object]. Bermúdez, however, disagrees: although he concedes that such studies cannot be explained without attributing to the infants content-bearing states, he denies that the data require attributing to them any concepts. On his view, an appropriate theory of concepts must acknowledge that possessing a specific concept involves engaging in certain proprietary inferences. Thus, he identifies the ability to enter into inferences as distinctive of states with conceptual content, as opposed to states with non-conceptual content. And he denies that we should explain infants' understanding of objects in terms of their inferential capacities. According to him, "the fact that young infants are surprised when certain expectations they have [about the behavior of physical objects] are thwarted does not mean that they have drawn inferences from the fact that something seems to be an example of an object" (69). Instead, the infant's behavior can be explained entirely in terms of their "parsing their visual array into bounded segments that (more or less) correspond to objects" (*ibid.*).

Regardless of the plausibility of this strategy with regards to earlier violation-of-

²⁵See Spelke & Van de Walle (1993) for one of the main studies that Bermúdez (1998) considers. Also, see Baillargeon *et al.* (2009) for a theoretical review of studies testing infants' physical knowledge.

expectation studies, it cannot account for the Scott and Baillargeon study.²⁶ For in the false-belief condition, the infants could not just *see* that the agent believed that the penguin under the transparent box was the one-piece penguin, let alone *see* that the agent would therefore reach for the opaque box. The reason is that it *wasn't* the one-piece penguin under the transparent box, and that even if it had been, the two penguins were perceptually identical. Thus, the results cannot be explained in terms of how the infants “parsed their visual array.” Rather, the infants had to recognize that the agent would incorrectly infer that the penguin under the transparent box was the one-piece penguin from the facts about how the penguins had appeared previously, and then recognize that from there the agent would infer that the two-piece penguin was be under the opaque box. Thus, the infants’ behavior cannot be accounted for without attributing to them the ability to infer from the fact that the agent had formed an initial false belief to the fact and that she would thereby come to have a second false belief. Not only did the infants draw inferences, they drew inferences *about the agent’s inferences*. For that reason, the notion of non-conceptual content as Bermúdez and others have primarily developed it is not adequate to account for their behavior.

Explaining the infants’ capacity for mindreading in terms of non-conceptual content will thus require explicating the notion of non-conceptual content in a manner that includes in its extension representations beyond those that figure in perceptual experience. But this is problematic as well, since the newly included representations will exhibit a number of the features that many philosophers have considered to be hallmarks of conceptual representation. For instance, one feature that many take to be distinctive of conceptual representation is a certain

²⁶For arguments against the view that infants’ object representations are solely perceptual see Carey (2009, Ch.2).

degree of *abstraction*. Thus, Weiskopf (*forthcoming*) argues that a necessary feature of conceptual thought is that it “involve[s] representing and categorizing the environment in ways that transcend the distinctions provided by a creature’s perceptual and motor systems.” But, as I just argued, the main reason that Bermúdez’s notion of non-conceptual content is inadequate for explaining infants’ PA mindreading is that the infants behavior requires attributing to them representations of others’ psychological states, and thus requires attributing to them representations with content that is too abstract to be wholly accounted for by appealing to perceptually available properties.

Another feature that many have taken to be necessary for a representation’s being a concept is its being under *endogenous control* of the organism, so that it can be used in psychological processes that are “decoupled” from the organism’s immediate spatiotemporal environment. As Prinz (2002) explains, “concepts are spontaneous: they are representations that can come under the endogenous control of the organism. Mere perceptual representations are receptive: they are controlled solely by exogenous stimuli... [C]oncepts can be freely deployed in thinking [but] perceptual representations are under environmental control” (197).²⁷ And although one may suspect that the infants’ capacities for PA mindreading are controlled solely by the exogenous stimuli of their immediate spatiotemporal circumstances, there is empirical evidence that suggests otherwise.

For instance, in an attempt to nullify concerns about attributing meta-PA’s to infants solely on the basis of their passively gazing at scenes with which they are presented, Buttelmann *et al.* (2009) employed an “active helping” paradigm. By taking advantage of the fact that infants

²⁷Also, see Sterelny (2003) on “decoupled representation” and Camp (2009a) on “stimulus-independence.”

over the age of 12 months have a strong tendency to help others achieve their goals, Buttelmann *et al.* tested whether 18-month-old infants could modulate their helping tendencies in a way that required an understanding of another's beliefs. The infant subjects were first trained how to open boxes by flipping a latch. Then they watched as a toy was switched from one box to another either while an agent also watched the switching (true-belief condition) or while the agent was absent (false-belief condition). Next, they watched as the agent attempted to open the box in which the toy was originally located but failed because of an inability to flip the latch. Finally, the infants were there encouraged to help the agent.

The results were as follows. In the true-belief condition, the infants got up from where they were sitting, navigated across the room, and opened the box the agent was trying to open (even though the toy was in the other box). In contrast, in the false-belief condition, they infants again actively helped the agent, but did so by opening the box where the toy was *actually* located. This suggests that in the true-belief condition, the infants recognized that the agent was aware of the true location of the toy and thus inferred that the agent must have wanted to open the first box for a reason other than retrieving the toy, and that in the false-belief condition, the infants recognized that the agent had a false belief regarding the location of the toy and thus wanted the toy but was looking for it where he falsely believed it to be. It appears, then, that in both cases the infants modulated their helping in accordance with what they took the agent to believe to be the toy's location. When conjoined with the results from Scott and Baillargeon (2009), which demonstrate that the representations implicated in infants' PA mindreading can be stored in long-term memory and recalled when necessary, these results strongly suggest that

infants' meta-PA's are under their endogenous control.²⁸ Admittedly, infants' capacity for exercising such control pales in comparison to the sort that adults exercise over their thought processes. But that is because their capacity is still developing, rather than because it does not exist.²⁹

These considerations force a dilemma upon the theorist who claims that infants' PA mindreading can be explained without ascribing to the infants the relevant concepts. On the one hand, she could try to explain the results using the notion of non-conceptual content as it has typically been developed. But this seems implausible since that notion of non-conceptual content is primarily meant to explain features of perceptual experience, and the infants were representing properties of their environment that transcend what is perceptually available. On the other hand, the theorist could explicate a more ecumenical notion of non-conceptual content that enables her to countenance the representations involved in infants' PA mindreading and the sorts of perceptual representations that have concerned many theorists of non-conceptual content as both

²⁸I take this to be consistent with the claim that aspects of PA mindreading are mandatory or automatic, in the sense that their operation is not under voluntary control. Thus I do not take it to be inconsistent with modularity-based accounts of mindreading that are committed to that claim (e.g., Scholl and Leslie 1999). An infant may not be able to control when the modular aspects of her mindreading capacity engage, yet still have control over how she reasons with the "output" of the module(s). In particular, mandatory modular mindreading systems could be (at least partly) responsible for the *generation* of candidate meta-PA's even though meta-PA *fixation* is under endogenous control and thus non-mandatory.

²⁹A property that is related to both entering into inferences and being under endogenous control and that many theorists understand as constitutive of conceptual representation is captured by what Evans (1982) calls the "Generality Constraint" (GC). According to GC, if a thinker can be legitimately said to think the thought *Fa*, and thus possess the concepts [*F*] and [*a*], she must have the capacity to think the thought that *Ga* and *Fb*. Unfortunately, because of space I cannot consider in detail how GC bears upon infants' capacity for PA mindreading. That said, as Carruthers (2009) points out, GC admits of interpretations of varying strength, depending upon the quantifier involved. So whether you think that the infants possess the relevant concepts will turn on which interpretation of GC you take to be appropriate. If GC just requires that a thinker who can think *Fa* can think *Ga* and *Fb* for *some G* and *some b* that she can think of, then, given the evidence currently available, one can make a compelling case that infants' PA mindreading abilities satisfy GC. If, however, GC requires that a thinker who can think *Fa* can think *Ga* and *Fb* for *every G* and *every b* that she can think of, things are not as clear. But, it should be stressed that the issue is wholly empirical, and that since there is now evidence that prelinguistic infants exhibit significant flexibility in the mindreading capacities, the burden of proof is arguably on the skeptic who thinks that infants are capable of forming representations that for some reason are unable to serve as constituents of their meta-PA's.

being instances of a single unified kind of content-bearing state. But doing so would ignore the many dissimilarities between the two types of representations, and would thus make the extension of the notion of non-conceptual content too heterogeneous to explain the commonalities amongst the various types of representation that fall under it. Moreover, such a taxonomy would disregard the many similarities shared between the representations involved in infants' PA mindreading and the representations that are typically thought to be concepts, including in particular their capacity to enter into inferences, their abstractness, and their being under endogenous control. The theorist who defends the claim that infants' PA mindreading ought to be explained without ascribing to the infants the relevant concepts because concept possession is impossible for non-linguistic creatures thus runs the risk of trivializing her position. If she wants to deny infants the capacity for conceptual PA mindreading, she will have to identify a property of adult PA mindreading that is lacking in infant PA mindreading and the presence of which in adults is inherently tied to their possessing a language. But, to avoid tautologies, it will have to be something other than the fact that adults have words to express the PA's generated by the psychological processes responsible for their mindreading. And, as I hope I have conveyed, given the sophistication of infant PA mindreading, this will be more difficult than it may initially seem.

1.4 CONCLUSION

Much interdisciplinary research is based upon the assumption that it is at least in principle possible for non-linguistic organisms to engage in PA mindreading. Bermúdez has recently challenged this assumption by trying to establish a necessary link between PA

mindreading and possessing a natural language. Here I have tried to neutralize Bermúdez's challenge in two ways. First, I argued that Bermúdez's argument in support of the necessary link is fallacious. Second, I supplemented my argument with empirical results that I argued indicate that prelinguistic infants engage in PA mindreading. When taken together, I think that these considerations undermine Bermúdez's claim that non-linguistic organisms are ipso facto incapable of PA mindreading.

This, however, raises the question of whether we ought to describe non-linguistic organisms that exhibit the capacity for PA mindreading as possessing the relevant concepts. Although the issue remains open, I have suggested that we should not think we must answer that question negatively because such organisms lack a natural language. Explaining infants' success on the non-verbal false-belief tasks requires attributing to them representations that both lack the features often identified as constitutive of states with non-conceptual content, and exhibit many features that are often considered to be distinctive of conceptual representation. This suggests that it is also in principle possible that non-human organisms possess the concepts involved in substantive PA mindreading, and that future research should continue to explore that possibility.

CHAPTER 2: MINDREADING, SIMULATION, AND INHIBITORY CONTROL

In the last chapter I argued in response to Bermúdez that the capacity for PA mindreading does not depend upon natural language possession. Upon arriving at this conclusion, the question that immediately arises is: what sort of cognitive processes *does* PA mindreading depend upon? What *are* the underlying neuropsychological mechanisms that allow mindreaders to attribute PA's to others? In this chapter I therefore want to shift the focus to considering how the recent evidence of infant PA mindreading bears on positive accounts of the cognitive architecture of mindreading, attending to the debate between “simulation theorists” and “theory theorists” in particular.³⁰

According to “theory-theory” (TT), our capacity for mindreading is based upon our (largely implicit or tacit) knowledge of various principles of psychology, knowledge that enables us to engage in factual reasoning about the psychological states of others. And, as the name suggests, defenders of this view conceive of the total sum of the principles that mindreaders in employ in mindreading as a “theory of mind” or “folk psychology.” In contrast, defenders of simulation theory (ST) “find it doubtful that the folk mentally possess or deploy such law[s] or generalization[s]” in their reasoning about others’ psychological states (Goldman 2006, 19). Instead, they argue, mindreaders can use their own minds as “simulations” of the minds of others. As Robert Gordon (2009), a leading defender of ST, explains, “the basic idea is that if the resources our own brain uses to guide our own behavior can be modified to work as representations of other people, then we have no need to store general information about what makes people tick: We just do the ticking for them” (1).

³⁰For overviews of this debate see Carruthers and Smith (1996), Davies, M. and Stone, T. (1995; 1995b) and Dougherty (2009).

Alvin Goldman (2006; 2009; 2009b) has recently driven the debate between theory-theorists and simulation-theorists significantly forward by articulating what is certainly the most thoroughly developed, empirically supported version of ST.³¹ In so doing, he has both raised the standards for what a comprehensive theory of mindreading should aim to accomplish, and clarified what it will take for an opponent of ST to call it into question. As Goldman explains, the early iterations of the debate focused on whether or not simulation- or theory-based processes played *any* role in mindreading. But amongst current participants in the debate there is general consensus that a hybrid account that acknowledges a role for both theory- and simulation-based cognitive processes in mindreading has the best chances of being correct. For instance, despite conceding that mindreading can occasionally proceed by simulation, Peter Carruthers (2006b) remains a vocal critic of ST because, like Goldman, he insists that the “focus should...be on the relative centrality of theory versus simulation in human mind-reading” (1). At this stage of the debate, then, the simulationist can no longer refute the defender of TT by showing that simulation plays *some* role in mindreading. Instead, as Goldman explains, “to gain victory over its rivals, [a defender of] ST must establish the *intensive* use of simulation [in mindreading] as defined by the *positive* characteristics” of simulative processes.³² Conversely, in order to refute ST, an opponent must show that such simulative processes do *not* play such an intensive role.

Here I will take up exactly that task, arguing that recent evidence from developmental psychology reveals that simulative processes as defined by Goldman *cannot* be responsible for what both sides of the debate agreed agree are some of the central phenomena to be explained by

³¹ For that reason I will assume that Goldman’s account is representative of the best that ST has to offer and set aside other versions to focus specifically on his.

³² Goldman (2006, 46; original emphasis). All further Goldman citations will refer to Goldman (2006) unless otherwise noted.

any satisfactory theory of mindreading. In doing so, I will concede to the general consensus that it is possible that simulation sometimes plays a role in mindreading. Thus, I will not attempt to demonstrate that simulative processes *never* factor in our attributing psychological states to others. Nor will I attempt to put forth a positive alternative to ST or defend TT. But, I will argue that simulation-based processes play a much more restricted role in mindreading than Goldman and other defenders of ST suggest they do. Whether or not my argument will therefore count as showing that simulation does not play an “intensive” role in mindreading—and thus whether or not it will constitute a refutation of ST—is a matter of debate. But, either way, it will certainly detract substantially from the overall plausibility of ST.

I will begin by explicating Goldman’s ST in §2.1, emphasizing how it relies critically on the claim that accurate simulationist PA mindreading requires the mindreader to have sufficiently developed the capacities for what Goldman calls “E-imagination” and “inhibitory control.” Then, in §2.2, I will explain how this feature of Goldman’s account enables him to provide a straightforward explanation of the developmental transition that occurs between ages three and four that allows four-year-olds to first successfully pass what is considered to be the litmus test of PA mindreading, the standard false-belief task: according to Goldman, it is between the ages of three and four that children first develop the inhibitory control (IC) necessary for engaging in PA mindreading. Importantly, I will concur with Goldman on this point. Indeed, I will argue that recent research from developmental psychology strongly substantiates both the claim that IC and PA mindreading are closely related psychological capacities, and the claim that the development of IC is what explains children’s transition to successful performance on standard false-belief tasks between ages three and four.

But, therein lies the problem for Goldman. According to ST there can be no accurate PA mindreading in the absence of IC. So when ST is conjoined with an IC account of the transition in children's performance on standard false-belief tasks between the ages of three and four, it predicts that we should find no evidence of children younger than approximately age three engaging in accurate PA mindreading. And, as I will reiterate in §2.3, such evidence has been found: the recent studies demonstrating that very young children—even infants as young as 13 mos.—can pass non-verbal false-belief tasks indicate clearly that accurate PA mindreading occurs well before children first develop the capacity for IC. There is, therefore, a three-way inconsistency between Goldman's ST, an IC account of standard false-belief task performance, and infant PA mindreading. Given the strength of the evidence both for infant mindreading and for the claim that the development of inhibitory control explains the transition in children's standard false-belief task performance, the most plausible conclusion to draw is that Goldman's ST is false.

2.1 GOLDMAN'S SIMULATION THEORY

The central contention of ST is that mindreading is dependent upon the mindreader using her own domain-general cognitive capacities—without recourse to any domain-specific reasoning mechanisms or tacit knowledge of a “theory of mind”—to discern others' psychological states. As Goldman elaborates, “a fundamental idea of ST is that mindreaders capitalize on the fact that they themselves are decision makers, hence possessors of decision-making capacities. To read the mind of others, they need not consult a special chapter on human psychology, containing a theory about the human decision-making mechanism” (20). Instead,

“because they have one of those mechanisms themselves, they can simply *run* their mechanism on the pretend input appropriate to the target’s initial position... In other words, mindreaders use their own minds to ‘mirror’ or ‘mimic’ the minds of others” (*ibid.*).

From this perspective, mindreading is in a number of ways analogous to what an aeronautical engineer does to predict how a particular plane design will perform in various aerodynamic environments. First, in both cases the predictor uses a model in certain specified target conditions to run a simulation of the process that is to be predicted. Moreover, in both cases the simulation generates an *accurate* prediction only to the extent that the predictor uses a model that *resembles* the process that is to be predicted. In the case of the engineer, the prediction is accurate because the process of a model plane flying in a wind tunnel resembles the process of an actual plane flying. In the case of mindreading, the prediction is accurate because the psychological processes used by the mindreader resemble the psychological processes of the mindreadee. Finally, in both cases the predictor’s use of the model counts as a “simulation” because, according to Goldman, a process P simulates another process P' just in case P “duplicates, replicates, or resembles P' in some significant respects (significant relative to the purposes or function of the task), and in its (significant) duplication of P' , P fulfills one of its purposes or functions” (37).³³

Goldman insists, however, that not all simulation-based mindreading is created equal. In

³³As Goldman (2006, p. 35) emphasizes, the notion of simulation at work in debates about mindreading is therefore best understood as distinct from the notion of simulation at work in computer science. In computer science, simulations do not necessarily work according to the same principles or undergo the same states or processes as that which is being simulated. For instance, the algorithms that govern a computer simulation of protein synthesis will often involve numerical calculations, but protein synthesis itself is not a calculative process. In contrast, a crucial aspect of the notion of simulation as used in debates about mindreading is that the mindreader *does* use the same – psychological processes and undergo the same psychological states as those she is trying predict.

particular, there is a distinction between “high-level” and “low-level” simulationist mindreading. The definitive distinction between the two types of mindreading involves the different kinds of neuropsychological mechanisms that underlie them.³⁴ More specifically, *low-level* mindreading is defined as involving “mirroring” processes, which are psychological processes “in which a neuron or neural system is activated as a result of observing another individual’s behavioral output or expression that results from, or would normally result from, an activation of a corresponding neuron or neural system in the observed individual” (2009b, 484). Alternatively, *high-level* mindreading is defined as involving “enactment imagination” (or “E-imagination”), which is an endogenously activated form of pretense that involves generating “mental surrogates” of the psychological states of the mindreadee (2006, 149).³⁵ According to Goldman, because E-imagination and mirroring processes are wholly distinct kinds of neuropsychological mechanisms, low-level and high-level mindreading are mutually exclusive: no instance of mindreading could simultaneously count as both.

Goldman also claims that the distinction between high-level and low-level mindreading roughly maps onto a distinction between the different kinds of psychological states being attributed: low-level mirroring processes are typically involved when the mindreader is attributing sensorimotor states and emotions, and high-level E-imaginative processes are

³⁴I say definitive because instead of appealing to underlying mechanisms, Goldman (2006) previously formulated the distinction between high-level and low-level mindreading in terms of the prototypical properties of each: high-level mindreading is typically subject to voluntary control, consciously accessible, and aimed at target states that are comparatively complex (e.g., PA’s), whereas low-level mindreading is typically psychologically primitive, automatic (as opposed to subject to voluntary control), largely inaccessible to consciousness, aimed at target states that are comparatively simple (e.g., sensations). However, de Vignemont (2009) criticized this formulation for being too elusive, since the mentioned prototypical properties are neither necessary nor sufficient for a psychological process to count as one or the other kind of mindreading. In response, Goldman (2009b) sharpened the distinction by appealing to the underlying psychological mechanisms.

³⁵Strictly speaking, the capacity for E-imagination can be used for purposes other than mindreading. But those alternative functions of E-imagination can be ignored for the time being.

typically involved when the mindreader is attributing PA's. This is not to deny that it is possible that some instances of PA mindreading (are at least partially) brought about by mirroring processes, nor is it to deny that one can generate predictions about another's sensorimotor states through the use of E-imagination. But, Goldman claims, our minds in general do not work that way: mirror systems are predominantly implicated in sensorimotor mindreading, and E-imagination is predominantly implicated in PA mindreading. Yet, the problem I want to raise for Goldman's ST specifically involves his account of PA mindreading. So, I will set Goldman's account of low-level mindreading aside in what follows and, instead, focus specifically on his E-imagination-based account of high-level PA mindreading.

E-imagination is not in and of itself sufficient for genuine PA mindreading. Instead, other cognitive mechanisms are necessary. In fact, on Goldman's account, the process of E-imagination-based PA mindreading requires four component stages. First, the mindreader *E-imagines* the appropriate first-order PA's. Second, she *simulates* those E-imagined PA's, which means that the E-imagined PA's serve as inputs to her decision-making or reasoning mechanisms, and that those mechanisms then process those inputs in order to generate an E-imagined state as output (i.e., an E-imagined decision, belief, desire, etc.). For this to work so that the resultant output does not issue in its typical causal effects, the E-imagined states must be processed while the mindreader's decision-making mechanisms are "offline" and the outputs of simulations must themselves be E-imaginative states. Otherwise, every instance of attempted mindreading would lead to a genuine decision (or belief, desire, etc.) and thereby cause the mindreader to actually act (or think, want, etc.) in accordance with the states she E-imagines. Third, a self-monitoring mechanism must *classify* the output of the simulation by identifying the

resultant first-order—and still E-imagined—PA's in terms of their mental state kind (e.g., belief, desire, etc.) and content (e.g., that the cat is on the mat, that it is raining, etc.). Fourth, on the basis of the classification of the outputs of the simulation, the mindreader *projects* the appropriate meta-PA's onto the mindreadee. Once the sequence reaches the projection stage, the mindreader is no longer engaging in either simulation or pretense. Instead, projection involves the mindreader forming a genuine, *non-E-imagined* meta-PA (i.e., a belief that another has a certain PA).

Accurate simulation-based PA mindreading, therefore, requires E-imagination, simulation, recognition, and projection. But, since my criticism of Goldman's ST mainly deals with his account of the relationship between the E-imagination and simulation stages, it will now be important to further explicate the details of those two stages. That said, it is crucial to recognize that on Goldman's account the mindreader does not actually form a meta-PA (i.e., does not actually represent the mindreadee as having a specific PA) until the projection stage of the mindreading process. The account therefore predicts that if an individual is wholly lacking any of the capacities required for earlier stages of the mindreading sequence, that individual should be wholly incapable of accurate PA mindreading. Likewise, it predicts that an individual should have a deficit in accurate PA mindreading to the degree that the individual has a deficit in any of the capacities required for the E-imagination, simulation, and classification stages of the mindreading process.

Now, as mentioned, Goldman considers E-imagination to be a kind of pretense. The idea here is that when one E-imagines a particular mental state, one undergoes a psychological experience that functionally and neurally replicates the experience one has when one actually has

the corresponding genuine (i.e., non-E-imagined) mental state. In other words, E-imagination involves “creating or trying to create in one’s own mind a selected mental state, or at least a rough facsimile of that state, through the faculty of imagination” (2006b, 42).³⁶

And by appealing to this distinctive E-imaginative form of pretense, Goldman (2006, 53) recognizes that he is, in effect, appealing to a novel theoretical construct. But he argues that doing so is warranted, invoking as support research on visual imagery and motor imagery, each of which he considers to be a type of E-imagination. For instance, when subjects hear descriptions of visual scenes while their eyes are closed, their saccade patterns closely resemble the saccade patterns that occur when subjects are actually presented with the visual scene that is depicted by those descriptions. And multiple studies have shown that subjects can use visual imagery to solve tasks at rates approximating those that the subjects would exhibit if presented with the actual visual scene (Shepard and Metzler, 1971; Shepard and Cooper, 1982). Similarly, Yue and Cole (1992) compared the results of just *motorically imagining* lifting weights and *actually* lifting weights over a four-week period. Predictably, they found that actually lifting weights increased the amount of force the subjects could generate while lifting by 30 percent. But, they also found that merely imagining lifting weights for the same time period actually increased the amount of force subjects could generate by 22 percent.

These results are, according to Goldman, most plausibly explained in terms of subjects

³⁶Thus, for Goldman E-imaginative pretense is not the same as what he calls “suppositional” pretense. This reason is that “merely supposing...doesn’t re-create the psychological circumstances operative” when one has the corresponding genuine psychological state. By contrast, replicating those psychological circumstances is *required* for simulation (175). To illustrate the difference, Goldman provides the example of trying to determine the visual appeal of a meal when considering whether to purchase the requisite ingredients from the grocer. In such a situation, in order to determine whether one would enjoy the relevant visual experience, one does not merely *suppose* that a plate with the ingredients on it is in front of oneself and infer what follows. Rather, Goldman claims, one E-imagines the plate to be in front of oneself, where this is done by generating in oneself a psychological state that resembles what one considers the visual experience of having the plate in front of oneself to be like.

spontaneously E-imagining the appropriate scenes (in the case of visual imagery) or movements (in the case of motor imagery). So, he argues, when conjoined with evidence from neuropsychology suggesting that the neural regions that underlie visual and motor imagery overlap significantly with the neural regions that underlie actual vision and motor behavior, they indicate that “visual [and motor] imagination is very much an enactment (simulation) of seeing [and moving] in behavioral as well as purely cognitive terms” (153). Moreover, he hypothesizes, just as we have the capacity to E-imagine perceptual and motor states, we have the capacity for a “conceptual” type of E-imagination that allows us to E-imagine PA’s in a manner that functionally and neurally closely replicates what it would be like to genuinely have those PA’s. And that it is precisely this conceptual E-imagination that he claims enables us to engage in PA mindreading.

Once again, though, E-imagination is not all that is required. Recall that after the mindreader has E-imagined the requisite PA’s, she must simulate them by inputting them into her offline practical-reasoning or decision-making mechanisms. Moreover, in order for the process of E-imagination-based mindreading to be of any use to the mindreader, it must enable her to attribute the *correct* psychological states to the mindreadee. So, it is essential to the accuracy of the mindreading process that the mindreader both E-imagine and simulate only the requisite PA’s. The mindreader must therefore both to be able to differentiate between genuine and E-imagined PA’s, and to govern which serve as input to the simulation routine. As Goldman explains, “it is...important to the success of a simulation for the attributor to *quarantine* [her] own idiosyncratic desires or beliefs (etc.) from the simulation routine. If the attributor has desires or beliefs that aren’t shared by the target, allowing them to seep into the routine could

contaminate it” (29; original emphasis). And such contamination would invariably lead to incorrect attributions during the projection stage of mindreading. Thus, on Goldman’s account, without the capacity to appropriately quarantine one’s own PA’s during the E-imagination and simulation stage of high-level mindreading, mindreaders will systematically err in their attempts to accurately attribute PA’s to others.

Goldman hypothesizes that the capacity to appropriately quarantine one’s own PA’s involves what cognitive scientists call “inhibitory control” (IC), which he describes as the cognitive ability to override a “prepotent” or dominant tendency to provide an incorrect response in order to select a response that is correct but inconsistent with the prepotent response. One example of how this ability is measured is the classic Stroop task, in which subjects are presented with a series of words, each of which is the name of a color but also is printed in a color that is different from the color it names (e.g., the word “red” printed in blue ink). In order to pass such tasks, subjects have to inhibit the prepotent tendency to respond based upon the words that they read (e.g. saying “red”) and instead respond based upon the color of the letters of the words that they read (e.g., saying “blue”). Another example, which is used in developmental studies, is the “Day-Night” task, where subjects are presented with a series of cards, each of which has on it either a picture of a sun or a picture of a moon. When given this task, in order to pass, subjects have to say “night” when they see the sun and “day” when they see the moon. And this requires inhibiting the prepotent tendency to respond by saying “day” when they see the sun and “night” when they see the moon (Gerstadt et al., 1994).³⁷

³⁷Tests of IC typically fall into one of two categories. The Stroop and “Day-Night” tasks are examples of “Conflict” tasks, which require subjects to not only refrain from impulsively responding incorrectly but also to provide a response that is inconsistent with the impulsive or prepotent response. In contrast, “Delay” tasks require subjects to

Given the function that Goldman hypothesizes the ability for inhibitory control performs in enabling the mindreader to appropriately quarantine her own PA's, he must predict that to the extent that an individual is lacking in the capacity for inhibitory control, she should be incapable of accurate mindreading. As he himself says, "if simulation is a major facet of third-person mindreading, successful mindreading should involve inhibition of the self-perspective. As a corollary, someone impaired in the ability to inhibit the self-perspective should have trouble producing accurate mindreading" (170). In other words, if Goldman's ST is correct, there should be no accurate high-level simulation without successful quarantining, and, as a result, there should be no accurate PA mindreading without inhibitory control.

2.2 SIMULATION THEORY AND INHIBITORY CONTROL

The essential role that Goldman affords IC in the process of PA mindreading is, as he sees things, an asset of his version of ST, since the theory can thereby explain straight-forwardly the psychological link between IC and PA mindreading that is suggested by a number of the more prominent patterns in the results generated by empirical studies of mindreading. In this respect, Goldman argues, ST is superior to TT, which he claims leaves the link between the two capacities poorly explained at best. It is, however, precisely the fact that Goldman makes PA mindreading as inextricably linked to IC as he does that I will argue undermines his account.

only delay, temper, or completely suppress an impulsive response. For instance, in the "Gift Delay" task, children are asked to refrain from peeking while an experimenter loudly wraps them a present (Kochanska et al. (1996). (Gerstadt et al. 1994). Children's performance on both types of tasks improves substantially during their preschool years, particularly between the ages of 3 and 4. However, as will become important in §3.3, children find "Conflict" tasks much more difficult than "Delay" tasks. See Carlson (2005) for an overview of measures of IC. For individual studies see Carlson and Moses, 2001 Diamond, 2002; Gerstadt *et al.*, 1994; Kochanska *et al.*, 1997, 2001; Zelazo *et al.*, 2003.

Importantly, though, I will not challenge his claim that IC and PA mindreading are closely related. Instead, I will only take issue with the precise role in PA mindreading that Goldman is forced to ascribe to IC because of his more general commitments to ST. Thus, before presenting my argument against Goldman's account, it will be critical to explain why I think he is correct to appeal to IC in attempting to account for what the relevant empirical studies suggest are some of the more salient features of our capacity for PA mindreading.

The principal pattern in the empirical results that Goldman is able to readily explain by making IC a necessary precondition for successful PA mindreading is the drastic improvement in children's performance on standard false-belief tasks (SFBT's) that occurs roughly between the ages of three and four. Recall from Chapter 2 that in such tasks, subjects typically watch an agent place an object in one location and then, while the agent is gone, they witness the object being moved to a new location. Afterwards, the subjects are asked to point or say where the agent will look for the object when she returns. Also recall that at age three children typically fail this task, answering that the agent will look where the object is actually located, but that by the time they are approximately four they consistently pass it, answering that the agent will look where she (falsely) believes the object is located.³⁸

Goldman explains the discrepancy between the two age groups as follows:

Given a false-belief scenario, a simulator must simulate the target with a pretend belief that contravenes what he knows to be true. He must use this feigned belief rather than his genuine one to predict (or retrodict) the target's belief. In a switched location task, for example, he must use a pretend belief that the object is in location L, when he knows that it's really in location L*...[H]e must quarantine or inhibit his genuine belief to keep it from infecting his simulation...[T]he transition from to 3 to 5 (or 4) years is marked by an enhanced capacity to do this job.³⁹

³⁸See Callaghan et al. (2005) for a review of evidence for the universality of this developmental timetable.

³⁹Goldman (2006, 197). In the original text, the last sentence of this passage actually begins "Perhaps the transition

On this view, three-year-olds do not have sufficiently developed IC capacities. So, when they are asked how the agent will behave, they are not be able to appropriately inhibit their own belief about the object's actual location from playing a role in—or “contaminating”—their simulative processes. As a result, their simulation takes as input an inappropriate set of PA's and they end up answering (incorrectly) that the agent will look for the object in its new, actual location. Four-year-olds, on the other hand, *do* have sufficiently developed IC capacities, and thus *are* able to appropriately inhibit their own non-E-imagined beliefs. As a result, their simulation takes as input only the appropriate E-imagined PA's, allowing them to (correctly) answer that the agent will look for the object in its previous location.

I will call any account that maintains that the development of IC explains, or at least partially explains, the transition to successful performance on SFBT's an “IC account” of SFBT performance. As Goldman emphasizes, such an account is corroborated—at least in broad outline—by multiple lines of evidence. First, a number of studies have shown that decreasing the IC demands placed on subjects significantly improves their performance on SFBT's and other related PA mindreading tasks. For instance, Wellman and Bartsch (1988) discovered that a number of manipulations that reduced the IC demands of SFBT's enabled 3-year-olds to perform at the same level that 4-year-olds usually do, including, in particular, both (i) explicitly telling the subject the agent's belief, and (ii) noting the inconsistency between the subject's belief and the agent's belief.⁴⁰ Similarly, Koós *et al.* (1997) found that when the perceptual salience of the

from...”, but the context indicates that Goldman intends to actually—and not just “perhaps”—endorse and defend the account expressed by the sentence, especially because he goes on to present a good portion of the data I mention in this section as supporting it. See also pp. 72-76, 164-173.

⁴⁰Interestingly, in all conditions where 3-year-olds performed as well as 4-year-olds, subjects were not told, and thus

object that the agent is looking for is decreased because the object is eaten by another character (and thereby removed from the scenario), children pass the task at age three. And when Zaitchik (1991) compared conditions where the perceptual salience of the object was reduced because subjects only knew its location through testimony and thus could not see it with conditions where the object was visible, she discovered that children found the decreased salience conditions significantly easier, and that 3-year-olds were able to successfully complete the task.⁴¹ Moreover, in their meta-analysis of nearly all previous studies that employed SFBT's, Wellman et al. (2001) found that in general experimental manipulations that reduced the IC demands of the task (either by decreasing the salience of the prepotent response or by increasing the salience of the agent's mental states) significantly improved performance. Interestingly, the same effect did not hold for manipulations that decreased other executive function demands (e.g., working memory or attentional demands).

Relatedly, it has been found that 3-year-olds are typically quite bad at deception, pointing to a reward's actual location when it is necessary to point to where the reward is *not* located in order to appropriately deceive an experimenter, and that children four and older typically succeed at such tasks (Sodian 1991; Russell et al. 1991). Carlson et al. (1998) hypothesized that this is because when confronted with these tasks children have a prepotent tendency to respond by pointing to the object's actual location (due to the fact that veridical pointing is a particularly

did not know, the actual location of the object. Moreover, in the one condition where subjects did know the object's true location, 3-year-olds failed the task, four-year-olds were at chance, and 4.5-year-olds passed. In other words, in the condition where subjects had to inhibit a response based upon their own knowledge of the object's location in order to correctly respond based upon the agent's belief, the discrepancy between subjects below the age of four and subjects above the age of four was still observed. This further supports the claim that successful performance on SFBT's requires the emergence of IC capacities.

⁴¹For similar results also see Freeman *et al.* (1991), Mitchell and Lacohee (1991), Moses (1993), and Robinson and Mitchell (1995).

well-practiced, habitual response), and found that when 3-year-olds were provided alternative means for deceiving the experimenter (e.g., using pictorial cues or gesturing with large arrows), they passed the deception tasks quite easily. Moreover, subjects' performance improved even more when they were able to deceive by merely hiding an object—and therefore did not have to respond in a way that directly contrasted with the object's actual location—instead deceiving by gesturing away from the object. Because “full fledged deception implicates an understanding of false belief [since] the very point of deception is to create a false belief in the mind of another for some ulterior purpose,” Carlson et al. took these results as evidence that decreased IC demands lead to improved performance on PA mindreading tasks (673).

A second line of support for the idea that the development of IC is responsible for improved SFBT performance comes from the work of researchers who have designed PA mindreading tasks with *increased* IC demands and found that subjects find them more difficult than control tasks. For instance, Leslie and Polizzi (1998) conducted a study using a variant of the SFBT in which the agent whose behavior was to be predicted had an “avoidance desire” (i.e., a desire for one of two objects that does *not* have some property). They argued that attributing an avoidance desire requires IC since it involves first identifying the object that does have the property to be avoided, thereby increasing the perceptual salience of that object, and then inferring that the other object lacks that property. Intriguingly, when the agent had an avoidance desire and a *true* belief, children's performance was nearly identical to their performance on SFBT's. But when the agent had an avoidance desire and a *false* belief, subjects' performance dropped significantly, with even 4- and 6-year-olds consistently failing to answer correctly. Leslie and Polizzi argued that this demonstrates it is not just the avoidance desire that causes

problems for subjects, but rather the cumulative IC demands of processing a false belief *and* an avoidance desire.⁴² Relatedly, German and Hehman (2006) compared PA mindreading in young and elderly adults, and found that even though the performance of both groups decreased when the IC demands of SFBT's were increased, the elderly adults' performance did so to a significantly greater degree. Moreover, the elderly were not impaired in PA mindreading in general; rather, their performance only suffered in conditions with high IC demands. German and Hehman argue that this is best explained in terms of the independently confirmed claim that elderly adults suffer from substantially diminished IC capacities.

A third line of evidence indicates that IC and PA mindreading are diachronically related as well. And although the exact implications of this evidence remain unclear, it certainly supports the claim that sufficiently developed IC is necessary for successful SFBT performance. To begin, children exhibit significant advances in their capacity for IC during the time period when they first begin to pass SFBT's (Carlson, 2005; Diamond and Taylor, 1996; Frye *et al.*, 1995). For example, Frye *et al.* (1995) asked children to repeatedly sort cards according to one rule and then suddenly switch to sorting the cards according to another rule. They found that 3-year-olds were significantly more likely to persevere in sorting according to the first rule than 4-year-olds and 5-year-olds. Similarly, Jones, Rothbart and Posner (2003) found that 3-year-olds consistently fail "Simon Says" tasks whereas 3.5-year-olds consistently pass them.⁴³ Since such

⁴²More recently, Leslie, German and Polizzi (2005) replicated these results, showing that although 4-year-olds generally pass tasks that require inhibiting one prepotent response, they significantly struggle with tasks that require simultaneously inhibiting multiple prepotent responses. For similar results, see Cassidy (1998) and Yadzi *et al.* (2006).

⁴³Moreover, as Carlson and Moses (2001) note, the behavioral evidence that IC goes through substantial development very close to the age that children begin to pass SFBT's is nicely corroborated by research on brain maturation. For instance, the frontal lobes are thought to be heavily involved in IC processing, and there is a sharp increase in frontal lobe maturation rate between beginning around age 4 (Luria, 1973; Thatcher, 1992).

tasks are generally assumed to be reliable measures of IC, these results provide further support for the claim that it is the emergence of IC that enables 4-year-olds but not 3-year-olds to pass SFBT's.

Other studies demonstrate that there is a strong within individual correlation between performance on IC and PA mindreading tasks. Carlson and Moses (2001) presented 3-year-olds and 4-year-olds with batteries of both IC and PA mindreading tasks. They first found that IC and PA mindreading scores were significantly correlated with age, such that 4-year-olds had much better IC and PA mindreading capacities. They also found that individual performance on IC tasks significantly correlated with performance on PA mindreading tasks, and that the correlation remained significant even after controlling for age, gender, verbal ability, general executive function abilities, performance on pretense tasks, and number of siblings. Closer analysis also revealed that "Conflict" IC tasks, which are similar to PA mindreading tasks in that they "require children to [simultaneously] inhibit a dominant response... as well as activate an incompatible, novel response," independently predicted PA mindreading performance. By contrast, "Delay" tasks, which only require suppressing a prepotent response for a certain time period, did not.⁴⁴ Similarly, Hughes (1998a) analyzed the relationship between executive function capacities in general and PA mindreading, and found that within individuals, early executive function performance at age four significantly predicted performance on PA mindreading tasks one year later even though there was no evidence of a reciprocal relation of PA mindreading at age four performance predicting IC performance at age 5.⁴⁵ Interestingly, closer analysis also revealed

⁴⁴See fn. 37 above for a description of "Conflict" and "Delay" tasks.

⁴⁵Once again the correlation remained after exact age, verbal ability and initial mindreading scores were factored out.

that the only individual executive function test that independently predicted later PA mindreading was a task that specifically measured IC skills.

More recently, Flynn (2007) has confirmed these results. With a group of subjects with a mean age of 3 years, 10 months, Flynn tested children's capacities for IC and PA mindreading over four weeks for six phases. The study replicated Hughes (1998) findings that early (i.e., phase 1) IC predicted later (i.e., phase 6) PA mindreading—again even after controlling for age, verbal ability and early PA mindreading performance—and that no such predictive relation held between early PA mindreading and later IC. This is especially interesting because Flynn's study was only 20 weeks, less than half the duration of Hughes' (1998) study, which means “IC makes a significant impact on [PA mindreading] development over months rather than years” (64). Moreover, Flynn's analyses of individual subjects showed that for the vast majority of them IC performance increased at either the same time or shortly before PA mindreading performance increased.⁴⁶

Lastly, Birch and Bloom (2003) demonstrated that when faced with “knowledge-attribution” tasks, children, like adults, are subject to the “curse of knowledge”—that is, a tendency to be biased towards one's own epistemic perspective when considering what others do

⁴⁶Other studies have found similar correlations. For instance, Carlson et al. (2004) found a predictive correlation between executive function skills, including IC capacities, at 24 mos. and mindreading at age 39 mos. And the correlation held after controlling for age, sex, verbal ability, maternal education and early mindreading. Similarly, Hughes and Ensor (2007) measured children at age 2, 3 and 4 on a battery of executive function and PA mindreading tasks. Importantly, the executive function battery included a large number of tasks that specifically tested IC. They found that in general early executive function performance predicted later PA mindreading performance (once again controlling for age and verbal, but also socioeconomic status), even though the reverse predictive relation did not hold. And cross-cultural studies shed further light on the diachronic relationship between IC and PA mindreading. Sabbagh et al. (2006) found that individual differences in IC task performance correlate strongly with SFBT performance in Chinese and U.S. children. Yet although 3.5-year-old Chinese children were advanced in IC relative to US children, they were not advanced in FB. Oh & Lewis (2008) confirmed this with Korean children. On the whole,, these results strongly suggest that IC is necessary but not sufficient for PA mindreading.

and do not know. In the knowledge-attribution tasks, children were told that an agent was either familiar or unfamiliar with a set of toys, and then told whether or not each toy contained an object inside it. However, half of the children were also actually shown the contents of each toy, whereas the other half were not. Next, the children were asked whether or not the agent knew what was inside the toy. Predictably, when the children were told that the agent was familiar with the toys, they were more likely to attribute knowledge to the agent. However, when subjects had actually seen the contents of the toys and thus knew what each contained, they were also more likely to attribute knowledge to the agent. In particular, when the agent was unfamiliar with the toys, children were significantly more likely to fail to inhibit their knowledge of the toys contents, causing them to mistakenly attribute knowledge to the agent knowledge. Most important for present concerns, Birch and Bloom also found that 3-year-olds, and to a lesser degree 4-year-olds, were significantly more susceptible to the failure of inhibition than 5-year-olds, and argued that this is plausibly due to the younger children's lack of sufficiently developed IC.

In sum, then, there is extensive evidence that IC and PA mindreading are closely linked psychologically. In general, the evidence suggests that children's IC capacities go through a major developmental transition between the ages of three and four, and that it is their undergoing that transition that allows them to first pass SFBT's. Just as Goldman himself says, "advances in inhibitory control" during the mentioned time period appear to constitute a "crucial enabling factor" for successful performance on tasks aimed at testing the children's capacity for PA mindreading (198).

2.3 MINDREADING WITHOUT INHIBITORY CONTROL

With the evidence supporting an IC account of SFBT performance in place, I am now in a position to put forth my argument against Goldman's ST. Recall that according to ST there should be no accurate simulation without IC, since in the absence of successful inhibition the mindreader's simulation will be contaminated by her own genuine, un-quarantined PA's. Also, recall that it is this feature of ST that Goldman argues readily accounts for the link between IC and PA mindreading suggested by the aforementioned developmental evidence: on his view, it is the developmental transition in children's IC capacities that *explains* the successful SFBT performance that begins around age four, since it is only then that children's IC capacities are sufficiently developed enough to allow for accurate, "uncontaminated" PA mindreading. From this it follows that Goldman is committed to the claim that children should be incapable of accurate PA mindreading prior to the age at which they begin passing SFBT's. As a result, his version of ST predicts that there can be no accurate PA mindreading before the time children are approximately 3 years old.

Yet, this prediction has been shown to be false by the studies employing non-verbal false belief tasks that I introduced in Chapter 2. Indeed, to date, evidence of PA mindreading from non-verbal tasks has been found with subjects that are 13 mos. (Sperber et al., 2007), 15 mos. (Onishi and Baillargeon, 2005), 18 mos. (Scott and Baillargeon, 2009), 25 mos. (Southgate et al., 2007), 2 yrs. 11 mos. to 3 yrs. 7 mos. (Clements and Perner, 1994), and 4 yrs. (Ruffman et al., 2001).⁴⁷ When considered together, these results provide compelling support for the claim that children's capacity for PA mindreading is present from shortly after their first birthday and on—

⁴⁷See also Garnham & Ruffman (2001) and Low (2010).

well before the age when children begin to pass SFBT's.

So, there is a three-way inconsistency between Goldman's ST account of high-level PA mindreading, the recent evidence of infant PA mindreading, and an IC account of SFBT performance. Any two could be true, but not all three at once. To the extent that the evidence for infant PA mindreading is taken as veridical, and to the extent that we also have strong empirical support for an IC account of standard false-belief task performance, the appropriate thing to do when confronted with the inconsistency is to conclude that Goldman's ST is a mistaken theory of PA mindreading.

Now upon being confronting with this argument, someone sympathetic to Goldman's ST may challenge the claim—entailed by my argument—that the infants that pass non-verbal false-belief tasks lack the IC capacities necessary for accurate PA mindreading. Such an objector may invoke for support the study by Buttelmann et al. (2009), pointing out that it involved a quite complicated “active-helping” non-verbal false-belief task, and that what was particularly remarkable about the results was the unexpected flexibility with which children as young as 18 mos. were able to employ their capacity for PA mindreading in the service of achieving their goal of assisting the agent. And, the objector may ask, assuming that such goal-direct behavior requires some degree of IC, don't such results provide at least *prima facie* support for the claim that infants *do* have the requisite IC for PA mindreading? If they do, there would be no inconsistency between Goldman's ST and infants passing non-verbal false-belief tasks. Given the independent plausibility of Goldman's ST, it may therefore be more appropriate to deny that infants lack the IC necessary for PA mindreading than to conclude, as I do, that Goldman's account is false.

I doubt, however, that this objection can be sustained. To see why, it is important to distinguish two ways in which it could be developed. On the one hand, the objection could be understood as claiming that the infants who pass non-verbal false-belief tasks have already *completely* developed the IC capacities necessary for accurate PA mindreading. On this view, even though infants' IC capacities may continue to develop as the infants get older, they have already developed past the minimum threshold required for accurate PA mindreading by shortly after the infants' first birthday. On the other hand, the objection could be understood as making the more nuanced claim that infants who pass non-verbal false-beliefs tasks have IC capacities that are developed *to the degree required to pass non-verbal false-belief tasks*—and thus to engage in accurate PA mindreading to some extent—even though they are not developed to the degree required to pass SFBT's (or other PA mindreading tasks with high IC demands). On this view, the inference from the premise that infants do not have IC capacities sufficient for them to pass SFBT's, to the conclusion that the infants do not have IC capacities sufficient for PA mindreading is, at best, defeasible, and requires further support if it is to substantiate the conclusion that Goldman's ST is false.

The first version of the objection really is indefensible. The problem is that it entails that some factor other than IC is *wholly* responsible for the transition in children's SFBT performance between the ages of three and four, since if infants have already *entirely* developed the IC necessary to pass SFBT's, there must be some *other* reason for their not passing them until age four. And the claim that development in IC is not even partially—let alone overwhelmingly or entirely—responsible for the transition in SFBT performance is implausible for a host of reasons.

First and foremost, as I showed in §3.2, the support for an IC account of SFBT

performance is quite extensive (and growing). So, in order for the claim that infants already have all the IC required to pass SFBT's to be taken seriously, deflationary explanations of the empirical results that provide that support would have to be supplied. In particular, the results of the numerous studies (Diamond, 2002; Diamond and Taylor, 1996; Gerstadt et al., 1994; Kochanska et al., 2001; Zelazo et al., 2003) that have shown, independent of any concern with PA mindreading, that children's IC capacities go through a substantial developmental transition at exactly the age that children begin to pass SFBT's would have to be explained away entirely. Given the strength and abundance of those results—let alone the results that actually indicate a relationship between IC and PA mindreading—I am reluctant to think that the requisite deflationary explanations could actually be compellingly formulated.

Second, if Goldman were to relinquish his adherence to an IC account of the developmental transition in SFBT performance then, at least from the focused perspective of my argument against his version of ST, he would accrue significant explanatory debt. Primarily, this is because he has claimed that it is an explanatory virtue of ST that it coheres so well with an IC account, so severing the connection between the two would both seriously detract from its purported superiority over its competitors, and necessitate a comprehensive retooling of the theory. Moreover, Goldman has explicitly defended such an account in the past, so for him to deny it now without independent reasons would be *ad hoc*.

Although the second version of the objection is more plausible than the first, I think that it is plagued by a pair of problems that show that it is untenable as well. The primary problem is that there is no independent positive evidence for the claim that children have developed the sort

of IC that Goldman has described as necessary for PA mindreading by the age at which they pass non-verbal false-belief tasks. To see why, note that to date all studies of IC in young children have employed tasks that can be grouped into one of two kinds: simple or complex.⁴⁸ In *simple* response inhibition tasks, subjects are only required to suppress a prepotent or automatic response tendency. For instance, in “delay of gratification” tasks, subjects are measured in how long they are able to inhibit a rewarding behavior (e.g., eating a candy, looking to see the gift the experimenter is wrapping, etc.). By contrast, in *complex* response inhibition tasks, subjects must both suppress a prepotent but incorrect response tendency *and* select instead a response that is correct but inconsistent with the prepotent response. For instance, in the “Grass-Snow” task, which is a variant of the Stroop task, subjects must point to a white button when the experimenter says “grass” and point to a green button when she says “snow.” But, according to Goldman, accurate PA mindreading requires one to both “E-imagine [and simulate and project] states of the target that differ from one’s own [and] quarantine states of one’s own that differ from the targets” (170). The reason is that merely quarantining one’s own conflicting PA’s is insufficient for accurate simulation-based PA mindreading; the accurate mindreader must instead quarantine her own conflicting PA’s *at the same time* as she is proceeding through the four stages of mindreading that Goldman’s ST describes. So, if Goldman is correct, infants could not pass non-verbal false-belief tasks simply by inhibiting a prepotent but incorrect response—that is, they could not pass simply by quarantining their own PA’s. Rather, they would need the capacity to simultaneously both suppress a prepotent and incorrect response (i.e., quarantine their own genuine PA’s) *and* select a correct response that is inconsistent with the prepotent response

⁴⁸This distinction closely maps onto the distinction between “Conflict” and “Delay” tasks mentioned above.

(i.e., E-imagine, simulate and project the appropriate but contrasting PA's). A child's satisfactorily completing only simple response inhibition tasks is therefore not sufficient for demonstrating that the child has the IC that Goldman's ST requires for PA mindreading. And, although there have been a number of studies that have tested infant's capacities for IC using simple response inhibition tasks, as far as I know, there currently exist no studies that have used complex response inhibition tasks with subjects younger than 22 mos., which is nearly one year later than the age at which they first pass non-verbal false-belief tasks. That being case, the objection that infants do in fact have the IC necessary for PA mindreading even if they lack the IC necessary for SFBT's is missing independent empirical support.

Of course, the current lack of positive evidence for the relevant kind of IC in infants at the ages at which they pass non-verbal false-belief tasks does not unequivocally undermine the claim that those infants do have IC; further studies could always uncover such evidence in the future. But, the paucity of empirical support for the objection is exacerbated by a second problem, which is that when one looks closely at the experimental methods of the studies that have generated the evidence of infant PA mindreading, one discovers that there are in fact no IC demands placed on the subjects. This point can be elucidated by comparing those methods to SFBT's.

In order for a subject to pass a SFBT at least the following three things are necessary. First, a mechanism for generating meta-PA's: there must be some process by which the mindreader is able to attribute PA's to the mindreadee. Second, a response-selection mechanism: when faced with the SFBT, the subject must be able to make use of their meta-PA in responding appropriately—that is, the subject must have appropriate cognitive access to the meta-PA. Third,

a conflicting-response inhibition mechanism: subjects must be able to override any prepotent tendency to respond based upon their own contrary epistemic perspective (i.e., respond based upon where they believe the object is located, rather than where they believe the agent falsely believes the object is located). Only if all three of these mechanisms are operative will the mindreader provide a correct response based upon the agent's false belief. Other factors can certainly play a role in SFBT performance, but without these three in particular subjects will systematically fail SFBT's.⁴⁹

By contrast, the tasks that have produced evidence of infant PA mindreading are not nearly as exacting. In particular, none of them require the inhibition of a conflicting response. Consider, for instance, the studies that employ “violation-of expectation” non-verbal false-belief tasks. In completing such tasks, infants look longer when presented with scenes where an agent acts in ways inconsistent with their expectations about how the agent will act. This is because in general infants spontaneously look significantly longer at scenes that violate their expectations—scenes, that is, that violate their beliefs about what will happen—than they do at scenes that are consistent with those expectations. Thus, in such tasks, subjects' responses are wholly caused by an inconsistency between what they observe and what they expect or believe will happen. But, when presented with scenes where an agent acts in ways inconsistent with subjects' expectations about how the agent will act given the agent's beliefs, subjects have no *other* expectations or beliefs that are inconsistent with what they observe. In particular, subjects' beliefs about where the object that the agent wants is *actually located* are not violated at all, since the object remains located where subjects believe it is throughout. Therefore, the subjects have no expectation other

⁴⁹For a similar analysis of the task demands of SFBT's see Baillargeon et al. (2010).

than the one they have about where the agent will look that could possibly serve as a response. In other words, the subjects that engage in such tasks have no prepotent conflicting response that needs to be inhibited and, consequently, they do not need any IC in order to do so successfully.

A similar analysis can be provided for the single study that has used an “anticipatory-looking” paradigm to demonstrate the existence of infant PA mindreading (Southgate et al., 2007). In contrast to violation-of-expectation tasks, in anticipatory-looking tasks subject’s responses are not governed by their disposition to spontaneously look significantly longer at scenes that violate their expectations. Instead, a number of factors play a role in determining subjects’ looking patterns, since a number of factors play a role in determining what they anticipate will happen. Thus, Southgate et al. knew that the subjects in their study would have various tendencies that interacted in complex ways to guide their responses. In particular, they predicted that subjects would have had a prepotent tendency to “look toward the box containing the ball, not necessarily because they expected the [agent] to search there, but because knowledge of the presence of the object and its significance in the situation might elicit saccades in that direction” (588).

For that exact reason, though, Southgate et al. designed their study so that the object for which the agent was looking was always removed from the scene prior to eliciting subjects’ responses. And, as predicted, the 2-year-old’s looking patterns indicated that their expectations of the agent’s behavior reflected a capacity for PA mindreading. But more important for present purposes, Southgate (personal communication) reported that in control conditions where the object was *not* removed, subjects failed the task by consistently looking towards the object’s actual location, rather than towards where the agent believed the object was located. Similarly,

Clements and Perner (1994) conducted a study using a comparable anticipatory-looking task, except that the object remained in the scenario for the duration of the task. And, although they found that the anticipatory gaze patterns of the subjects older than the age of three strongly implicated a capacity for PA mindreading, they also found that subjects younger than three systematically looked towards the object's actual location, thereby failing to correctly visually anticipate the behavior of an agent with a false belief.

On the whole, then, it appears that when the object for which the agent is looking is fully removed from the scenario, subjects are much less likely to form an explicit belief about the object's location that is inconsistent with the belief about the object's location that they attribute to the agent. Yet, if that is correct, there is good reason to think that the subjects in the Southgate et al. study did not have to override any prepotent tendency to respond based upon their own conflicting epistemic perspective: if they didn't form a belief about the object's location, they didn't have a belief to inhibit. Therefore, the study provides no evidence for the claim that infants that engage in PA mindreading also have IC.

Finally, consider the "active-helping" paradigm from the Buttleman et al. (2009) study that I suggested might initially seem to support the objection being considered. The purported reason why that study shows that the infants involved have the IC required for PA mindreading is the unprecedented flexibility with which they employed their capacity for PA mindreading. Earlier studies had been limited by the fact that they required drawing conclusions about infants' PA mindreading solely on the basis of infants' differential looking patterns. For that reason, they left the door open for the skeptical worry that such tasks could be solved by some low-level, non-mindreading process. And the active-helping paradigm mitigated that worry, since the infants

who correctly modulated their helping behavior on the basis of their beliefs about the agent's false beliefs demonstrated that they were able to use their PA mindreading as a means to helping the agent in the scenario achieve his goal—a much more cognitively sophisticated feat than simply gazing in the correct direction.

However, infants' cognitive sophistication notwithstanding, attending to these facts actually indicates why the task does *not* require any IC. In responding the way they did, the infants would have needed the capacity for IC only if they had to override a prepotent tendency to respond based upon their own psychological states that were inconsistent with the psychological states that they attributed to the agent. But, the best explanation of the infants' behavior suggests that their helping was most directly caused by (1) their altruistic desire to help the agent in the scenario achieve his goal and (2) their belief(s) about the agent's goal. IC therefore would have been required only if the infants simultaneously had either (3) a strong tendency to not help others achieve their goals or (4) a goal that directly conflicted with the agent's goal to retrieve the toy. Importantly, the infants did not need IC to inhibit their belief about the object's location, which in the false-belief condition *was* inconsistent with the belief that they attributed to the agent about the object's location. The reason is that unlike (1) and (2), the infant's belief about the agent's belief was not a direct cause of their helping. Instead, as Baillargeon et al. (2010) explain, “the infants...represented the agent's false belief and used this representation to infer what goal the agent was trying to achieve...[Thus] when prompted, the infants only needed to consult this additional information [viz., their belief about the agent's goal]: they did not have to tap their representation of the agent's false belief directly” (116). In other words, the infants' responses were not directly based upon their beliefs about the agent's

belief of the object's location. Consequently, they did not have to inhibit their own beliefs about the object's location in order to respond the way they did. So, attributing to infants the capacity for IC would only be warranted if there is also good reason to attribute to them either (3) or (4). But the whole study was based upon the premise that infants have the tendency *to* help others achieve their goals, which is the exact opposite (3), and there is no independent reason to attribute to them (4), since in general children have the exact opposite goal (i.e., the goal *to* retrieve toys). Thus, the Buttelmann et al. study provides no reason to attribute to the infants the capacity for IC.

Taken together, these considerations shift the burden back to the defender of ST: the objection that the infants that engage in PA mindreading do in fact have IC should only be taken serious at this point if it is coupled with (a) some reason to think that the relevant non-verbal false-belief tasks do, contrary to my analyses, impose IC demands on their subjects and (b) independent evidence that infants can meet those demands in non-mindreading control tasks. Otherwise, the fact that older children fail the exact sorts of IC tasks that one would predict them to pass if they had IC, and the fact that close inspection shows that non-verbal false-belief tasks do not impose any IC burdens each provide good reason to be skeptical of the claim that infants have IC. And, at least at this point, neither (a) or (b) exists. So, the defender of ST has no reason for claiming that infants that engage in PA mindreading do have IC other than they *must* in order for ST to be true. Accordingly, the evidence of infants passing non-verbal false-belief tasks strongly indicates the existence of PA mindreading without IC. But, unfortunately for Goldman, insofar there exists PA mindreading without inhibitory control ST is false.

2.4 CONCLUSION

I have argued that when conjoined with an IC account of SFBT performance, Goldman's ST predicts that children younger than approximately age three should be incapable of accurate mindreading. For that reason, though, Goldman's ST is inconsistent with a growing body of evidence that infants have the capacity for PA mindreading. This strongly suggests that Goldman's account is false. Moreover, I have argued that the response to my argument that is *prima facie* most plausible—the response that infants that pass non-verbal false-belief tasks have already developed the IC that Goldman claims is required for accurate PA mindreading—is, in the end, indefensible. This is both because there is no independent evidence that infants have the relevant sort of IC, and because none of the tasks that provide the evidence of infant PA mindreading require IC to pass.

Having reached this conclusion, it is worth reemphasizing that Goldman sharply distinguishes between low-level and high-level mindreading, and nothing I have claimed conflicts with Goldman's simulationist account of low-level mindreading. It is likewise worth reemphasizing that my claim is not that simulation-based processes *never* play a role in mindreading. As noted earlier, there is a near unanimous agreement amongst mindreading researchers that the capacity for mindreading is more heterogeneous than previously thought, and that this improves the prospects for a hybrid account that acknowledges the role of both theory- and simulation-based processes in mindreading. Thus, as Goldman himself emphasizes, the debate between defenders and opponents of ST should no longer be conceived of as a debate about whether or not simulative processes *do in fact* play a role in mindreading. Instead, the debate should be about the *degree* to which simulation-based and theory-based processes play a

role in mindreading. For both of these reasons, it remains entirely possible that simulation plays *some* role in our capacity for mindreading.

That being case, my argument nonetheless severely restricts the explanatory scope of ST, since children's performance on standard (and now non-verbal) false-belief tasks is one of the foremost explananda for which any adequate theory of mindreading must account. Thus, my argument undermines Goldman's project of showing that simulation plays an "intensive" role in PA mindreading. Whether or not this ought to count as a refutation of ST is, as noted, debatable. But it certainly shifts the burden of proof to the defenders of ST, making it incumbent upon them to show that simulation still plays a substantive role in PA mindreading. Moreover, the problem that I raise for Goldman is generated by his account of the cognitive processes required for the E-imagination and simulation stages of the PA mindreading process. But, unfortunately for his sake, Goldman (2006, 34) claims that the two critical positive features of ST that set it apart from alternative accounts of mindreading are the hypothesized role of pretend (viz., E-imagined) states and the hypothesis that the mindreader uses the very same mechanisms as the target (viz., that the mindreader uses a simulation). So, there is no manner in which ST can be revised to escape the problem I have raised for it while retaining its distinctiveness as a theory.

BIBLIOGRAPHY

Baillargeon, R. et al. (2010) False-belief understanding in infants. *Trends in Cognitive Science*.

Baillargeon, R. et al. (2009). A new account of infants physical reasoning. In Woodward, A. & Needham, A. (eds.) *Learning and the Infant Mind*, 66-116.

Baillargeon, R. (2004). Infants reasoning about hidden objects: evidence for event-general and event-specific expectations. *Developmental Science*, 7:4. 391-424.

Baillargeon, R. (2000). Reply to Bogartz, Shinskey, and Schilling; Schilling; and Cashon and Cohen. *Infancy*, 1:4, 447-462.

Baron-Cohen, S., Tager-Flusberg, H. & Cohen, D. (2000). *Understanding Other Minds*. New York: Oxford UP.

Beckoff, M., Allen, C., & Burghardt, G. (2002). *The Cognitive Animal*. Cambridge, MA: MIT Press.

Benson, J. & Sabbagh, M. (2010). Theory of Mind and Executive Functioning. In Zelazo, P.D., Chandler, M, & Crone, E. (eds.), *Developmental Social Cognitive Neuroscience*, Psychology Press: New York, NY.

Bermúdez, J. (forthcoming). Two Arguments for the Language Dependence of Conceptual Thought, *Grazer Philosophische Studien*.

Bermúdez, J. (2009). Mindreading in the animal kingdom? in Lurz, R. (ed.) *The Philosophy of Animal Minds*.

Bermúdez, J. (2007). What is at stake in the debate on nonconceptual content? *Philosophical Perspectives*, 21, 55-72.

Bermúdez, J. (2003). *Thinking Without Words*. New York: Oxford UP.

Bermúdez, J. (2000). Concepts and the Priority Principle. *Psychology*, 11, 33.

Bermúdez, J. (1998). *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press.

Birch, S. & Bloom, P. (2003) Children are cursed: An asymmetric bias in mental-state attribution. *Psychol. Sci.* 14, 283–286.

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

- Birch, S. (2005). When knowledge is a curse: Children's and adults' reasoning about mental states. *Current Directions in Psychological Science*, 14, 25–29.
- Birch, S. & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences*, 8, 255–260.
- Block, N. (2007). *Consciousness, Function and Representation*. Cambridge, MA: MIT Press.
- Bloom, P., & German, T.P. (2000). Two reasons to abandon the false-belief task as a test of theory of mind. *Cognition*, 77, B25–B31.
- Buttelmann, D. et al. (2009) Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112, 337–342.
- Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12:5, 187-192.
- Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., Tapanya, S., & Singh, S. (2005). Synchrony in the onset of mental state reasoning. *Psychological Science*, 16, 378–384.
- Camp, E. (2009a). Putting thoughts to work: Concepts, systematicity, and stimulus-independence. *Philosophy and Phenomenological Research*, 78:2, 275-311.
- Camp, E. (2009b). A Language of Baboon Thought? in Lurz, R. (ed.) *The Philosophy of Animal Minds*.
- Carlson, S., Mandell, D., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology*, 40, 1105–1122.
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, 28, 595–616.
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72, 1032–1053.
- Carlson, S. M., Moses, L. J., & 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. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development*, 69, 672–691.

- Carruthers, P. (2009). Invertebrate concepts confront the Generality Constraint (and win). in Lurz, R. (ed.) *The Philosophy of Animal Minds*.
- Carruthers, P. (2009b). How we know our own minds: the relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, 32, 164-176.
- Carruthers, P. (2009c). Simulation and the first person. *Philosophical Studies*.
- Carruthers, P. (2006). *The Architecture of Mind*. Oxford: Oxford UP.
- Carruthers, P. (2006b). Review of Simulation Minds. *NDPR*.
- Carruthers, P. (2005). *Consciousness: Essays from a Higher-Order Perspective*. Oxford: Oxford UP.
- Carruthers, P. (2004). On Being Simple-Minded. Ch. 12 in *Consciousness: Essays from a Higher-Order Perspective*.
- Carruthers, P. & Boucher, J. (1998). *Language and Thought: Interdisciplinary Themes*. Cambridge, UK: Cambridge UP.
- Carruthers, P. & Chamberlain, A. (2000). *Evolution and the Human Mind*. Cambridge, UK: Cambridge UP.
- Carruthers, P. & Smith, P. (1996). *Theories of Theories of Mind*. Cambridge, UK: Cambridge UP.
- Cassidy, K. W. (1998). Three- and four-year-old children's ability to use desire and belief-based reasoning. *Cognition*, 66, 1–11.
- Clark, A. (1998). Magic words: How language augments human computation. in Carruthers, P. & Boucher, J. (eds.) *Language and Thought: Interdisciplinary Themes*.
- Clark, A. (2009). *Supersizing the Mind*. Oxford: Oxford UP.
- Clements, W.A. & Perner, J. (1994) Implicit understanding of belief. *Cognitive Development*. 9, 377–395.
- Cohen, J. & McLaughlin, B. (2007) *Contemporary Debates in Philosophy of Mind*, Malden, MA: Blackwell Publishing.
- Davies, M. & Stone, T. (1995), *Folk Psychology: The Theory of Mind Debate*. Oxford:

Blackwell.

Davies, M. & Stone, T. (1995b), *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.

de Vignemont, F. (2009) Drawing the boundary between low-level and high-level mindreading. *Philosophical Studies*, 144:3, pp. 457-466.

de Villiers, J. (2000). Language and theory of mind: What are the developmental relationships? in Baron-Cohen, S., Tager-Flusberg, H. & Cohen, D. (eds.) *Understanding Other Minds*.

Dougherty, M.J. (2009). *Theory of Mind*. Psychology Press: New York, NY.

Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford UP.

Flynn, E. (2007). The role of inhibitory control in false belief understanding. *Infant and Child Development*, 16, 53–69.

Flynn, E. (2006). A microgenetic investigation of stability and continuity in theory of mind development. *British Journal of Developmental Psychology*, 24, 631–654.

Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science*, 7, 103–115.

Fodor, J. (1975). *The Language of Thought*. Cambridge, MA: Harvard UP.

Friedman, O., & Leslie, A. M. (2005). Processing demands in belief–desire reasoning: Inhibition or general difficulty? *Developmental Science*, 8, 218–225.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, 10, 483–527.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, 7, 77–83.

Garfield, J., Peterson, C., & Perry, T. (2001). Social cognition, language acquisition and the development of the theory of mind. *Mind & Language*, 16, 494-541.

Garnham, W.A. & Ruffman, T. (2001) Doesn't see, doesn't know: Is anticipatory looking really related to understanding of belief? *Developmental Science*, 4, 94–100

German, T. & Hehman, J.A. (2006). Representational and executive selection resources in

'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition*, 101, 129-152.

Goldman, A. (2006). *Simulating Minds*. Oxford University Press.

Goldman, A. (2006b). Imagination and Simulation in Audience Responses to Fiction. In S. Nichols (ed.), *The Architecture of Imagination*.

Goldman, A. I. (2008). Mirroring, mindreading, and simulation. In J. Pineda (Ed.), *Mirror neuronsystems: The role of mirroring processes in social cognition*. New York: Humana Press.

Goldman, A. (2009). Replies to the Perner and Brandl, Saxe, Vignemont, and Carruthers. *Philosophical Studies*, 144, 477-491.

Goldman, A. I. (2009b). Mirroring, simulating, and mindreading. *Mind and Language*, 24, 235–252.

Gordon, R. (2009) Folk Psychology as Mental Simulation. *Stanford Encyclopedia of Philosophy*.

Hala, S. et al. (2003) Executive functioning and false-belief understanding in preschool children: Two tasks are harder than one. *J. Cogn. Dev.* 4, 275–298.

Hauser, M. & Spaulding (2006). Wild rhesus monkeys generate causal inferences about possible and impossible physical transformations in the absence of experience. *PNAS*, 103: 18, 7181-7185.

Heck, R. (2007). Are there different kinds of content? in Cohen, J. & McLaughlin, B. (eds.) *Contemporary Debates in Philosophy of Mind*, 117-138.

Hughes, C. (1998a). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, 16, 233–253.

Hughes, C. (1998b). Finding your marbles: Does pre-schoolers' strategic behaviour predict later understanding of mind? *Developmental Psychology*, 34, 1326–1339.

Jones, L. B., Rothbart, M. K., & Posner, M. I. (2003). Development of executive attention in preschool children. *Developmental Science*, 6(5), 498-504.

Kain, W., & Perner, J. (2005). What fMRI can tell us about the ToM–EF connection: False beliefs, working memory, and inhibition. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability and theory of mind*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Kloo, D., & Perner, J. (2003). Training transfer between card sorting and false belief understanding: Helping children apply conflicting descriptions. *Child Development, 74*, 1823–1839.
- Koos, O., Gergely, G., Csibra, G., & Biro, S. (1997, April). Why eating Smarties makes you smart: Understanding of false belief at the age of 3. Poster presented at the biennial meeting of the Society for Research in Child Development, Washington, DC.
- Leslie, A. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfeld & S. Gelman (eds.), *Mapping the Mind*, Cambridge University Press.
- Leslie, A. & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science, 1*, 247-253.
- Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in “theory of mind.” *Trends in Cognitive Sciences, 8*, 529–533.
- Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief–desire reasoning as a process of selection. *Cognitive Psychology, 50*, 45–85.
- Liu, D. et al. (2008) Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Dev. Psychol. 44*, 523–531.
- Low, J. (2010) Preschoolers implicit and explicit false-belief understanding: Relations with complex syntactical mastery. *Child Development*.
- Lurz, R. (2009). *The Philosophy of Animal Minds*, Cambridge UP.
- Moses, L. J. (2001). Executive accounts of theory-of-mind development. *Child Development, 72*, 688–690.
- Moses, L. J., & Sabbagh, M. A. (2007). Interactions between domain general and domain specific processes in the development of children’s theories of mind. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition*. New York: Psychology Press.
- Müller, U., Zelazo, P. D., & Imrisek, S. (2004). Executive function and children’s understanding of false belief: How specific is the relation? *Cognitive Development, 20*, 173–189.
- Nichols, S. & Stich, S. (2003). *Mindreading*. Oxford University Press.
- Oh, S., & Lewis, C. (2008). Korean preschoolers’ advanced inhibitory control and its relation to

- other executive skills and mental state understanding. *Child Development*, 79, 80–99.
- Onishi, K. & Baillargeon, R. (2005). Do 15-month-olds understand false beliefs? *Science*, 308, 255-258.
- Onishi, K., Baillargeon, R., & Leslie, A. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica*, 124, 106-128.
- Papafragou, A., Cassidy, K. & Gleitman, L. (2007). When we think about thinking: the acquisition of belief verbs. *Cognition*, 105, 125-165.
- Penn, D. & Povinelli, D. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a theory of mind. *Phil. Trans. R. Soc. B.*, 362:1480, 731-744.
- Perner, J. (1998). The meta-intentional nature of executive functions and theory of mind. In P. Carruthers & J. Boucher (Eds.), *Language and thought* (pp. 270–283). Cambridge: Cambridge University Press.
- Perner, J. & Ruffman, T. (2005) Infants’ insight into the mind: How deep? *Science*, 308, 214–21.
- Povinelli, D. & Vonk, J. (2004). We don’t need a microscope to explore the chimpanzees mind. *Mind & Language*, 19, 1-28.
- Premack, D. G. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515-526.
- Ruffman, T. & Perner, J. (2005) Do infants really understand false belief?: Response to Leslie. *Trends in Cognitive Science*, 9, 462–463.
- Russell, J. (1996). *Agency. Its role in mental development*, Hove: Erlbaum (UK), Taylor & Francis Ltd.
- Russell, J. (1998). How executive disorders can bring about an inadequate ‘theory of mind’. In J. Russell (Ed.), *Autism as an executive disorder* (pp. 256–299), Oxford: Oxford University Press.
- Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science*, 17, 74–81.
- Scott, R. & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, 80, 1172-1196.
- Smith, P. (1996). Language and the evolution of mindreading. in Carruthers, P. & Smith, P.

(eds.) *Theories of Theories of Mind*.

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

Song, H., Onishi, K., Baillargeon, R., & 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., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18, 587-592.

Sterelny, K. (2003). *Thought in a Hostile World*. Malden, MA: Blackwell Publishing.

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

Tetzlaff, M. & Rey, G. (2009). Systematicity and intentional realism in honeybee navigation. in Lurz, R. (ed.) *The Philosophy of Animal Minds*.

Weiskopf, D. (forthcoming) Concepts and the modularity of thought. *Dialectica*.

Wellman, H., Cross, D. & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655-684.

Wellman, H., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, 30, 239-277.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development*, 72, 655-684.

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

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

Woodward, A. & Needham, A. (2009) *Learning and the Infant Mind*. New York: Oxford University Press.

Yazdi, A.A., German, T., Defeyter, M.A. & Siegal, M. (2006). Competence and performance in belief-desire reasoning across two cultures: The truth, the whole truth and nothing but the truth about false belief? *Cognition*, 100, 343-368.

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition*, 35, 41–69.

Zaitchik, D. (1991) Is only seeing really believing? Sources of the true belief in the false belief task. *Cognitive Development*, 6, 91-103.