December 2018

Seeing and Believing: Philosophical Issues in Theory of Mind Development

Joseph A. Hedger

Syracuse University

Abstract

All human beings understand the behaviors of others as causal results of their mental states. Philosophers call this ability *folk psychology* and developmental researchers call it *theory of mind (ToM)*. My dissertation concerns how this reasoning works and how it is acquired.

First, I develop and expand a theory of how folk psychology develops in childhood. This is the *Perceptual Access Reasoning, or PAR* theory of the Fabricius lab. Contrary to the two views dominant in the field, I argue that ToM (*belief reasoning* or BR) is acquired around 6 years of age after undergoing two preliminary cognitive stages, *reality reasoning* (RR) and *perceptual access reasoning* (PAR). Neither of the first two satges of ToM development involve an understanding of mental representation. Evidence for the PAR hypothesis comes from the failure of 4- and 5-year-olds on a false belief task which includes a third, irrelevant, alternative; their failure on true belief tasks; and their failure on no belief tasks. Only the PAR hypothesis can account for all the data. Chapter 2 explains the PAR hypothesis and children's understanding of *believing*. Chapter 3 extends the PAR theory to children's understanding of perception, and demonstrates that the data (mostly tasks testing Flavell's classic 4 levels model of perception understanding and his appearance/reality distinction) support the PAR hypothesis.

Second, I demonstrate how this theory can be usefully applied to solve problems in cognitive science. In Chapter 4 I explore dual systems theories of cognition (and ToM in particular). In Chapter 5 I solve the Perner-Povinelli Problem—the claim that no empirical test can decide whether subjects are using mentalist rules to pass ToM tasks, or merely using

behavioral rules which require no understanding of mental representation. In Chapter 6 I use the PAR hypothesis to argue that a limited theory-theory of concepts is plausible. The PAR stage concept of KNOWING and the adult (BR) concept of KNOWING are fundamentally different because the former is non-representational. Evidence for this is that children in the PAR stage do not distinguish between knowing and guessing correctly, nor between lying and being mistakenly incorrect. The PAR child's concept of KNOWING is inextricably linked with perceptual access and correct behavior; in other words, with the inferential rules of the PAR theory. I then defend this hypothesis against Fodor's shareability objection.

Finally, in Chapter 7, I make some specific suggestions for continuing my folk psychology research program by expanding the PAR theory and applying it to other problems in philosophy.

Seeing and Believing:

Philosophical Issues in Theory of Mind Development

By

Joseph A. Hedger

B.A., Arizona State University, 2007
M.A., Arizona State University, 2010

Dissertation

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Philosophy.

Syracuse University

December 2018

TABLE OF CONTENTS

LIST OF ILLUSTRATIVE MATERIALS

CHAPTER ONE

Introduction: Methodology and Overview

1. Methodology: Speculative Psychology

Jerry Fodor prefaces *The Language of Thought* (1975) by calling it a work of speculative psychology. He describes this discipline as follows:

> It wasn't quite philosophy because it was concerned with empirical theory construction. It wasn't quite psychology because it wasn't an experimental science. But it used the methods of both philosophy and psychology because it was dedicated to the notion that scientific theories should be both conceptually disciplined and empirically constrained.

The monograph you are holding is also a work of speculative psychology, and as such I take Fodor 1975 and Fodor 1983 to be models of the methodology that I shall employ.

There is a place for empirically-minded philosophers of mind to work alongside psychologists in ways that are mutually beneficial. Perhaps one of the lessons of the earliest research to come out of the recent Experimental Philosophy movement is that philosophers receive insufficient training in the getting-your-hands-dirty work of empirical data gathering. It's safe to say that controlling for confounding variables in a thought experiment is much easier than doing it in a lab or on a questionnaire. Psychologists have a lot of training and experience in the gathering and mathematical analysis of data. On the other hand, philosophers are trained theorists. They spend a lot of time coming up with theories and making sure that they're consistent with other commonly held beliefs. The best philosophers are also good at making sure that their theories are consistent with both science and common sense. There is a process of reflective equilibrium in which one balances broader theories with individual findings from experiments. Philosophers are also good at interpreting data and constructing theories that don't draw conclusions that go excessively beyond what the data suport. In this way philosophers and psychologists can work well together, the philosophers constraining the theorizing and the psychologists doing the empirical work, and both coming up with new ideas and designing future experiments.

Of course, in taking empirical psychology seriously as a source for theorizing, this methodology continues a tradition in philosophy of mind which began in the late 1960s and early 1970s (as far as I can tell). I see exemplars of this tradition in the work of e.g. Thomas Nagel (at least in "Brain Bisection and the Unity of Consciousness"), Ned Block, Tyler Burge, Robert Van Gulick, Daniel Dennett, C. L. Hardin, Eric Schwitzgebel, Georges Rey, Peter Carruthers, Jesse Prinz, Stephen Stich, and of course Jerry Fodor. Although the theories of these philosophers differ in many respects, they are all alike in taking empirical data from psychology experiments as both starting points and constraints when theorizing about the mind.[1]

2. Philosophers and Cognitive Development of Theory of Mind

[1] Here I am thinking of myself as belonging to the Analytic tradition in philosophy, because that's what my training is in. In taking the science of psychology seriously as both a constraint and a data source for theorizing, Continental philosophers were probably further ahead (thinking of for instance Merleau-Ponty's attitude toward child psychology).

All people, regardless of their culture, automatically interpret the behavior of other people by ascribing various mental states and personality traits to those people. It is so ubiquitous that it is almost unnoticeable, but I think it would be inconceivable to do otherwise, even if (as a few philosophers such as Stich and the Churchlands argue) this "folk theory" ends up being largely false. For instance, imagine you are walking down a familiar street when you notice that a person on a skateboard is heading straight for you. Contrast that situation with what you would expect if you noticed a skateboard heading straight for you with no person on it. Your predictions about the behavior of the skateboard are going to be drastically different in the distinct scenarios.[2] The prediction you apply in the case of the person is going to incorporate what you interpret from that person's facial expression about what she believes and perceives, along with attributing to her a general desire not to collide with anyone or anything. The other situation, by contrast, will only involve a rudimentary understanding of physics.

Philosophers call the theory of human behavior utilized in the first situation "folk psychology," but psychologists have been calling it "Theory of Mind" (ToM) since at least Premack & Woodruff (1978). The field's inception is a paradigmatic example of psychologists and philosophers working together fruitfully. In the original commentaries on psychologists David Premack and Guy Woodruff's chimpanzee studies (1978), a handful of philosophers suggested an experiment which was to become the *false belief task* (Bennet 1978, Dennett 1978, Harman 1978, Pylyshyn 1978). This experiment was adapted for testing children by

---

[2] This is, incidentally, one reason why I find Dennett's Intentional Stance view so lacking—it ignores crucial differences between systems with minds and systems without them (unless he wants to ascribe minds to lightning bolts and thermostats).

psychologists Heinz Wimmer and Josef Perner (1983). In a multitude of variations, and with subjects ranging from infants to adults, from chimps to dogs, crows and rats, this experiment has been the centerpiece of study on ToM and its development.

For almost four decades now, many psychologists have spent their time investigating how this theory of mind develops. There have been a few philosophers paying attention to this research program, including e.g. about a third of the papers in Carruthers & Smith (1996), the dissertation and early work of Eric Schwitzgebel (which received very little attention from the rest of the discipline) and numerous references in the work of Stephen Stich and Peter Carruthers. Most of this attention resulted in a few philosophers weighing in on the Theory-Theory/ Simulation Theory debate (e.g. Goldman 2006, Nichols & Stich 2003), but that formerly hot topic is no longer as popular, and most philosophers never went further than this rather narrow spectrum. However, there is whole host of philosophical puzzles and issues in the psychological literature which are waiting to be fruitfully addressed by philosophers using the Speculative Psychology method outlined above, and these have largely been ignored by philosophers. The present monograph is an attempt to improve that situation.

The time is ripe for a philosophical investigation of this sort. The field of ToM in Developmental Psychology has itself been in disarray during the past decade, as the recent finding that infants are able to pass the false belief task has sent shockwaves through university psychology departments and created upheaval amongst researchers who were content with the *Traditional View* that children undergo a profound cognitive shift at about the age of four when they putatively acquire the concept of belief (e.g. Gopnik & Wellman, 1992; Perner, 1988, 1991; Saxe, Carey, & Kanwisher, 2004; Wellman et al, 2001). On the other hand, the *Nativist View* is that ToM ability, including an understanding of belief, is innate.[3] Although this view has been around since the 1980s (Leslie 1987, see also Fodor 1992), it has gained prominence in the past decade following Onishi & Baillargeon (2005)'s landmark study demonstrating that 15-month-olds can pass a non-verbal false belief (location) task. A flurry of studies has followed since Onishi and Baillargeon (2005) first cracked the dam (Buttelman et al 2009, He et al 2011, Scott & Baillargeon 2009; Scott et al 2010, Song & Baillargeon 2008, Southgate et al 2010, Surian et al 2007, etc). So many studies have been done at this point in time that it would be implausible to deny that infants are capable of passing a non-verbal false task. This throws a major monkey wrench into the Traditional View, which had finally seemed to settle in as the status quo by the end of the second millennium.

Although the current theory of mind debate in psychology (with a few philosophers recently weighing in; see e.g. Carruthers 2013) is consumed with the Traditional versus the Nativist views, recent findings that 4- and 5-year-olds *fail* the true belief task (Fabricius et al 2010; discussed in Chapter 2) raise major problems for both sides. If children understand beliefs at age 4 as the Traditional View holds, then the true belief task should be passed at the same time as the false belief task. In fact, even though the task demands are as made as commensurable as possible, intuitively the true belief task ought to be easier to pass than the false belief task.

[3] See Chapter 2 below for fuller explanations of the Traditional and Nativist views of ToM development.

The true belief findings are, if anything, even worse for the Nativist View. According to this hypothesis the failure of 3-year-olds on false belief tasks is not a sign that they lack *competence* in using BR, but is simply a kind of *performance error* (Chomsky, 1965; Fodor, 1992) in which something interferes with this competence to cause them to fail false belief tasks. Thus, according to Nativism, failure of inhibition or something similar causes 3-year-olds to fail a verbal false belief task, but this is not an issue with looking time measures (Baillargeon et al 2010, Leslie 2005, Roth & Leslie 1998). In order to save the hypothesis, nativists now need an explanation for why 4- and 5-year-olds are making performance errors after they have apparently overcome the failure of inhibition in order to pass the false belief task. The Perceptual Access Reasoning (PAR) hypothesis of Fabricius and colleagues, unlike its more popular rivals, is able to explain all of these disparate findings, as I will demonstrate in the following chapter.

The present project continues to use the PAR hypothesis in order to deal with philosophical puzzles and difficulties in the development of Theory of Mind. I have already been involved in some fleshing out and development of this theory (Hedger & Fabricius 2011, Hedger 2016), and that continues in this monograph with a treatment of children's understanding of perception (Chapter 3) and belief (Chapter 2) from the PAR perspective. I also use the theory as a framework and springboard to solve problems such as the "Povinelli Problem" in Cognitive Science (Chapter 5), and in order to shed light on other murky topics such as concept acquisition (Chapter 6).

3. The Rest of the Monograph

Chapter 2 explains the Perceptual Access Reasoning (PAR) hypothesis, and then uses the resulting theory to explain young children's conception of belief, which progresses according to three distinct stages of development. I demonstrate that the PAR hypothesis is able to explain all of the data from developmental psychology, including the *prima facie* puzzling findings that infants and four-year-olds pass the false belief task but three-year-olds fail it, and that four- and five-year-olds fail the true belief task while three-year-olds and six-year-olds pass it. In this way the PAR hypothesis is shown to be superior to the two dominant views of the field—Nativism (that understanding of belief is innate and in fact is evident in infancy) and the Traditional View (that children develop belief understanding at about four years of age). In contrast, according to the PAR theory, genuine understanding of belief is not acquired until around six years of age. I also speculate that the PAR hypothesis is equipped to explain findings in the Theory of Mind of chimpanzees, adults, and persons with Autism Spectrum Disorder (ASD).

Chapter 3 examines empirical data concerning young children's understanding of perception, especially Flavell's influential model of perception understanding and the Level 1/ Level 2 distinction, and the work of Flavell and others on children's understanding of the appearance/ reality (A/R) distinction. If someone has the concept of belief, then they ought to be able to pass Level 2 and A/R tasks. However, as I argue in Chapter 3, someone using PAR should fail them. Initially, it appears that success rates of 4- and 5-year-olds vary widely on both sorts of tasks. However, some analysis and clarification demonstrates that many tasks commonly assumed to be tests of Level 2 understanding are actually testing only Level 1. Once these tasks are distinguished, we can see that 4- and 5-year-olds are seen to fail most Level 2 tasks, with the possible exception of some very simple ones. However, these latter are explained by the fact that understanding perception is a matter of degree, so that the same PAR subject may be able to pass very simple Level 2 tasks where the demands are not much higher than Level 1 tasks, but still perform poorly on slightly more complicated (and realistic) Level 2 tasks. I also argue that performance of 4- and 5-year-olds is consistent with what we would predict given their lack of understanding of mental states. Finally, I demonstrate that the PAR hypotheses is able to explain the findings during this transitional period of development that have so puzzled other researchers.

Chapter 4 explains the PAR hypothesis in relation to dual systems theories of cognition. I argue against an alternative hypothesis that is proposed by Anika Fiebich in a recent *Synthese* paper (2014). According to Fiebich, PAR is not a distinct transitional stage in children's theory of mind development, but is a fast and frugal System 1 heuristic that fades once children become fluent in social reasoning. I point out a number of problems with Fiebich's proposal and argue for the superiority of the PAR hypothesis. I also present five reasons to be skeptical about the findings of Perner and Horn, which purportedly show that 4- and 5-year-olds can pass the 3-location false belief task when the test is suitably modified. This is a further difficulty for Fiebich’s proposal, since she relies on these findings in her fluency theory. Finally, I sketch a dual systems theory of mind account based upon the PAR hypothesis which is different from Fiebich’s.

Chapter 5 addresses what has become known as the “Povinelli Problem” in cognitive science. Psychologists worry about finding an empirical method for determining whether chimps and infants are passing ToM tasks by using mentalist rules, which include a conception of mental representation, or whether they are passing them merely by using behavioral rules which require no understanding of the mind as such. Daniel Povinelli has made the argument that there is in fact no empirical method for deciding which of these two explanations is true, and Josef Perner has recently applied the argument to the case of human subjects. In this chapter, I demonstrate how the PAR theory can provide an answer to this challenge. In fact, we already have evidence that many 4- and 5-year-olds have been using rules which make no reference to beliefs in order to pass false belief tasks, and infants and chimps may be using these same behavioral rules (or some slight variation thereof). I provide an empirical test for determining whether or not this is indeed the case, thus solving the problem.

Chapter 6 tackles the thorny issue of concept acquisition. Implicit in the background of the PAR theory is the idea that the concept BELIEF is acquired at around age six, as a constituent of the Theory of Mind which they also develop at this time. This places it firmly in the tradition of Sue Carey's Theory-Theory of Concepts. However, while extremely popular amongst developmental psychologists (for reasons I make plain in this chapter), this view has been the victim of vehement attacks by Jerry Fodor. In this chapter I make explicit the PAR theory of concept acquisition which has been in the background in the previous chapters, and answer one of Fodor's challenges to the Theory-Theory view. I also explain that various theories which have gone under the moniker of "Theory-Theory" in psychology, which most people (including Carey and Fodor) have assumed are completely independent hypotheses, nonetheless actually mutually reinforce each other and provide a unified, coherent view of the human mind.

Finally, in Chapter 7 I sketch some fruitful directions and projects (both philosophical and empirical) based upon the PAR hypothesis, as part of a plan to extend the research program in this monograph into the future. Some projects further develop and expand the PAR theory, while others are applications to different philosophical problems. Finally, I provide a detailed sketch for my next project, viz. an exploration of the phenomenal experience of young children entitled *What is it like to be a 5-year-old?*

References

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

Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences, 1*, 557-560.

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

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

Carruthers, P., & Smith, P. K. (1996). *Theories of theories of mind.* Cambridge, England: Cambridge University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: The MIT Press.

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences, 1*, 568-570.

Fiebich, A. (2014). Mindreading with ease? Fluency and belief reasoning in 4- to 5-year-olds. *Synthese, 191,* 929-944.

Fodor, J. A. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

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

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition, 44,* 283-296.

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

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7,* 145-171.

Harman, G. (1978) Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences*

1, 576–77.

He, Z., Bolz, M., & Baillargeon, R. (2011). False-belief understanding in 2.5-year-olds: Evidence from violation-of expectation change-of-location and unexpected-contents tasks. *Developmental Science 14*, 292–305.

Hedger, J. A. (2016). Perceptual access reasoning: Developmental stage or System 1 heuristic? *Phenomenology and the Cognitive Sciences, 15,* 207-226.

Hedger, J. A., & Fabricius, W. V. (2011). True belief belies false belief: Recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development." *The Review of Philosophy and Psychology, 2*, 429-447.

Leslie, A. M. (1987). Pretense and representation: The origins of 'theory of mind.' *Psychological Review, 94,* 412-426.

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

Nichols, S., & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds.* Oxford, England: Oxford University Press.

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

Perner, J. (1988). Developing semantics for theories of mind: From propositional attitudes to mental representation. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 141-172). New York: Cambridge University Press.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: The MIT Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515-526.

Pylyshyn, Z. (1978). When is attribution of belief justified? *Behavioral and Brain Sciences, 1*, 592-593.

Roth, D., & Leslie, A. M. (1998). Solving belief problems: Toward a task analysis. *Cognition, 66,* 1-31.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55,* 87-124.

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

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

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

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

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

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.

CHAPTER TWO

Believing:

The Perceptual Access Reasoning hypothesis and young children's understanding of belief

1. Introduction

The study of children's understanding of mental states, along with the ability to use that information to predict the behavior of others, is called "Theory of Mind" (ToM) or "mindreading" in psychology. This chapter and the next will explore in detail the state of the art in children's ToM, focusing on the mental states of belief (this chapter) and perception (Chapter 3). The focus is for two main reasons: First, the study of these two has dominated the literature in psychology, and so the information available concerning these two types outweighs any others by a significant portion. Second, for what philosophers of mind call "folk psychology," these two types of states are seen to be the most crucial and prolific (see e.g. Churchland 1981, Davidson 1963, Dennett 1981, Fodor 1987, Horgan & Woodward 1985, Lewis 1972, Ramsey, Stich & Garon 1990). Pick your favorite toy example of folk psychology (Bob sees that it's raining and believes that an umbrella will keep him dry, so I he will go to the closet to get his umbrella before leaving the house) and perception and belief states are likely to appear in that explanation. Indeed, it's difficult to imagine *any* causal explanation of human behavior which takes into account the agent's mental states that could avoid including that agent's beliefs.[4]

[4] Williamson (1995, 2000) argues that knowing is the primary, more basic concept, and that belief is parasitic upon knowledge. On this view (which I am sympathetic to), some causal explanations of behavior might refer to knowing but not believing. Since this issue is tangential to the main points of this chapter, and won't really change anything

When do children understand the concept of belief? A massive amount of research in developmental psychology over the past 30 years has been devoted to this topic. However, most research has been narrowly focused on various versions of the false belief task. It has come to be taken for granted that this single task is the litmus test for Theory of Mind (ToM) ability.[5] However, Fabricius & Imbens-Bailey (2000) reveal a confound in the task allowing a child to give the correct answer either by attributing a false belief to the protagonist, or by attributing only ignorance to the protagonist and assuming that the protagonist will therefore be wrong about where to find the object. Furthermore, most researchers have assumed that the true belief task is too simple to be of much use in studying ToM development, but according to the PAR hypothesis children in an intermediate stage of mindreading ability should *fail* the true belief task, and Fabricius et al (2010) indeed confirmed that many 4- and 5-year-olds who pass the false belief task nonetheless fail the true belief task.

In this chapter, I focus on children's understanding of belief in order to explain the Perceptual Access Reasoning (PAR) hypothesis. This hypothesis is supported by numerous studies involving the false belief task (Section 2), the true belief task (Section 4), and the no belief task (Section 5). I then argue for its superiority over the two most popular views in theory of mind development, which I call the Traditional View and the Nativist View (Section 7).

---

of importance, I set this philosophical debate aside. See Section 3.1 of Chapter 7, below, for more about the theory of ToM development I argue for here and its relation to Williamson's view about knowledge.

[5] For example, Henry Wellman, one of the pioneers of the ToM field, and his colleagues recently write, "Our measure of theory of mind is a battery of false belief tasks. Explicit false-belief understanding is a milestone, universal theory-of-mind achievement of the preschool years (Wellman et al., 2001), and is the most commonly used measure in research examining individual differences in theory of mind during the preschool years..." (Wellman et al, 2011).

2. The false belief task

False belief tasks have been used as the core test of children's ability to attribute representational mental states and use them to understand behavior since the first study of theory of mind in children (Wimmer & Perner, 1983). In one such task, the *unexpected transfer task (or false location task)*, children are presented with the following scenario, normally accompanied by dolls to represent the characters in the story, and props to represent the locations: Maxi comes into the kitchen with a piece of chocolate. He wants to put it somewhere so he can find it when he comes back from playing outside. He puts it into the red cupboard, and then goes outside to play. Later, Maxi's mother comes into the kitchen to clean. She moves the chocolate from the red cupboard to the green cupboard while she is cleaning. Then, Maxi comes inside from playing. He is hungry and wants his chocolate. Subjects are then asked the test question: "Where will Maxi look for his chocolate?" The correct answer is, of course, the red cupboard, since Maxi ought to believe (falsely) that this is where his chocolate is located.

Philosopher Daniel Dennett (1978) first sketched the logic and rationale for false belief tasks.[6] Dennett was commenting on studies of theory of mind in chimpanzees (Premack & Woodruff, 1978), in which chimps would often predict that an agent who gave cues to his mental state (e.g., reaching for bananas hung from the ceiling) would subsequently engage in the correct behavior (i.e., get a box to stand on). Chimpanzees appeared to attribute mental

---

[6] Jonathon Bennet (1978), Gilbert Harman (1978), and Zenon Pylyshyn (1978) made similar suggestions in regards to studying chimpanzee theory of mind, and other philosophers had previously stressed the importance of understanding false belief in order to count as having the concept BELIEF. Donald Davidson (1975), e.g. argues that "Someone cannot have a belief unless he understands the possibility of being mistaken, and this requires grasping the contrast between truth and error—true belief and false belief" (p. 170).

states, but they might have simply associated familiar behaviors and outcomes. Dennett argued that an unambiguous test of mental state attribution required the chimpanzee to predict that the agent would engage in an *incorrect* behavior, which could be accomplished by giving the agent a false belief that the incorrect behavior would accomplish the goal. Wimmer and Perner (1983) adopted Dennett's logic to design the unexpected transfer task, which was the first false belief task for children. Thus the field began with a confidence that low-level mechanisms, such as might be used by chimpanzees, had been controlled for by using false belief tasks.

The *unexpected contents task (or Smarties task)* is a second false belief task (Hogrefe, Wimmer, & Perner, 1986). The child is shown a familiar candy container, such as an M&M bag, and asked what he thinks is inside. After the child says "M&M's," it is revealed that something unexpected is inside, such as a pencil. After the pencil is placed back inside the bag, the subject is informed that a friend of the experimenter's (named "Elmo" e.g.) is waiting outside. The test question is: "If he just looks at it, what will Elmo think is inside the bag?" The correct answer is to attribute to Elmo a false belief that M&M's are inside the bag.

Children under the age of 4 tend to fail verbal false belief tasks. In the examples above, they tend to choose the green cupboard, because that is where the chocolate currently is, despite the fact that Maxi was outside and could not have known that the chocolate was moved. They also say that Elmo will think that there is a pencil in the bag, because that is what the bag actually contains, even though Elmo has no reason to think that. Hedger & Fabricius (2011) refer to this as *Reality Reasoning (RR).* A meta-analysis of 178 false belief studies by

Wellman, Cross, and Watson (2001) revealed that the age at which 50% of children pass false belief tasks is 3 years, 8 months.

Importantly, the meta-analysis also revealed that none of the various ways researchers have used to try to make the tasks more understandable or motivating for younger children reliably eliminated RR and allowed younger children to pass at rates comparable to older children. Six variables were found to affect the age at which children passed false belief tasks, but no variation was able to improve the performance of 3 year, 5 month old subjects to chance or greater. Thus, Wellman et al (2001) argued that the consistent failure of younger children was due to a lack of conceptual understanding. Most psychologists and neuroscientists (e.g. Gopnik & Wellman, 1992; Perner, 1988, 1991; Saxe, Carey, & Kanwisher, 2004) have long agreed that at around 4 years of age children acquire the concept of representational mental states and transition from RR to what Fabricius & Hedger (2011) dub *Belief Reasoning (BR)*. BR signals a full-fledged theory of mind in which understanding and predicting the behavior of others is accomplished by attributing to them mental states, including false beliefs.

3. The PAR hypothesis

Dennett (1978) felt that true belief tasks would be too easy to pass and would not demonstrate anything of theoretical importance. That is true as long as there are only two ways that children could think about mental states – to not think about them at all (RR), or to correctly understand their representational nature (BR) – both of which predict success on true belief tasks. But when Fabricius and Imbens-Bailey (2000) proposed that children might be using what they called *Perceptual Access Reasoning (PAR)* instead of BR when they passed false belief tasks, it introduced an intermediate level of understanding and entailed the counterintuitive prediction that children should fail true belief tasks.

According to the PAR hypothesis (Fabricius & Imbens-Bailey 2000, Fabricius & Khalil 2003, Fabricius et al 2010, Hedger & Fabricius 2011, Hedger 2016), false belief tasks are confounded in a way that allows children to pass without truly understanding false beliefs. In the unexpected transfer task, for example, children might only reason that when Maxi returns he will not see his chocolate in the new location and will therefore not know it is there, and that as a consequence of his ignorance he will "get it wrong" and look in the empty location. But it just so happens that the one "wrong" location is also the location where children who attribute a false belief to Maxi should say he will look. Thus children can pass the task by attributing only ignorance, predicting unsuccessful behavior, and using a process of elimination. This line of reasoning does not involve attributing false belief. Likewise, children can pass the unexpected contents task by using PAR. They can reason that Elmo cannot see the pencil inside, so he will not know it is a pencil inside, so when given the forced choice (of a pencil or M&M's) he will "get it wrong" and say that M&M's are inside instead. PAR is thus composed of two rules:

Rule 1: Seeing → Knowing (and Not Seeing → Not Knowing)

Rule 2: Knowing → Getting it Right (and Not Knowing → Getting it Wrong)

Importantly, these two rules contain no reference to false beliefs (as such). In the first test of the PAR hypothesis, Fabricius and Kahlil (2003) modified all the various false belief tasks (versions of unexpected transfer, unexpected contents, and appearance-realty tasks)[7] by including a third, irrelevant alternative. Children using PAR should have no reason to choose the false belief alternative over the irrelevant alternative, but children using BR should. In three studies of over 150 5- and 6-year-olds, Fabricius and Kahlil (2003) found that, consistent with the PAR hypothesis, many children passed the standard tasks and failed the modified versions by choosing the irrelevant alternatives.[8]

The two PAR rules also contain no reference to true beliefs. The rules predict correct behavior on the basis of the child's decision that the other knows, but this does not presume that the child uses an adult conception of knowing as true, justified belief; namely, as a representational mental state that the person maintains over time and throughout changes in external reality.[9] Instead, knowing in PAR is caused by what the person has perceptual access to in the immediate situation, and both perceptual access and hence knowing change as the

---

[7] The unexpected transfer and unexpected contents tasks are described above. See Chapter 3, below, for the appearance-reality task.

[8] Of course, "many subjects passed the standard version but failed the modified version" is a quick gloss, because the paper involves 10 tasks and 3 studies. As a more specific example, in Study 1, an average of 70% of the 84 5-year-olds passed the standard false belief task, but an average of only 44% passed the modified 3-choice versions. See Fabricius and Khalil (2003) for further details.
Perner and Horn (2003) tested the PAR hypothesis using a variation of the three-alternative false belief procedure designed to be simpler than that used by Fabricius and Khalil, and they concluded that their findings refuted the hypothesis. However, Hedger (2016) provides a number of reasons to doubt these findings. Additionally, Gonzales et al (2013) failed to replicate Perner and Horn's findings against the hypothesis. See section 3.1 of Chapter 4 below for more details.

[9] Of course, 2600 years of Western Philosophy has failed to come up with an adequate analysis of knowledge, and the last thing I want to do is to attempt to define *knowing* here. The main point is that the PAR child's concept KNOW is non-representational. More on this in the chapters which follow, especially Chapter 5 and Chapter 6. Hence, even for philosophers who disagree with Williamson's view that knowledge is more basic than belief (1995, 2000), since a belief is a mental representation, but the PAR child's concept KNOW is non-representational, then there is no mention of "belief" (explicit or otherwise) in the PAR rules.

situation changes.[10] Thus PAR entails the prediction that children will judge that a protagonist will get it wrong regardless of whether the protagonist has a false belief or a true belief, as long as the situations in both cases result in comparable lack of perceptual access. To see this consider the various true belief versions of the false belief tasks (unexpected contents, unexpected transfer, and appearance-reality) used in Fabricius et al (2010).

4. The true belief task

The true belief version of the contents task begins like the false belief version. The child is shown an M&M bag and asked what he thinks is inside, and is then shown that the bag contains a pencil. It becomes a true belief task when the child watches the experimenter remove the pencil and fill the bag with M&M's before being asked the test question. Children using BR should reason that seeing the familiar candy container will cause Elmo to have a true belief that it contains M&M's. Children using RR should also answer M&M's, but simply because that is what the bag actually contains. Children using PAR should reason that Elmo cannot see the M&M's inside, so he will not know M&M's are inside, so when given the forced choice (of a pencil or M&M's) he will "get it wrong" and say that a pencil is inside. Using PAR, in other words, should cause children to fail the true belief version of the contents task just as it should cause them to pass the false belief version, because the protagonist is equally deprived of perceptual access to the contents in both cases.

[10] There is independent evidence that the child's understanding of 'know' undergoes much development during these years regarding its relation to both perceptual access and correct versus incorrect behavior (e.g., Perner 1991).

Fabricius et al (2010) used two slightly different true belief versions of the transfer task. In the first (used in Study 1), Maxi watches his mother place the chocolate in the red cupboard before leaving. During his absence, she takes it out and considers moving it to the green cupboard, but then changes her mind and returns the chocolate to its original location. Therefore this true belief transfer task is similar to the false belief transfer task in that the chocolate is moved during Maxi's absence, but it is not moved to a new location. In the second task (used in Study 2), Maxi watches his mother place his chocolate in the red cupboard, and then continues to watch as she changes her mind and moves it to the green cupboard. Then Maxi leaves and there is no subsequent movement of the chocolate. Hence, the chocolate is moved to a new location just as in the false belief version, but Maxi is present to witness the transfer. Thus, in both true belief tasks, when Maxi returns he should have a true belief that the chocolate is still in the green cupboard. Therefore children using BR should pass both versions of this task. But for children using PAR, when Maxi leaves his perceptual access to the situation is broken, and they do not attribute a belief to him that persists after he leaves the situation. When Maxi returns, children using PAR see him in a new situation, and when asked where Maxi will look, they should reason that since Maxi can't see the chocolate now, he doesn't know where it is, and thus he will "get it wrong" and look in the empty cupboard.

The true belief version of the appearance-reality task that Fabricius et al (2010) used (in Study 3) involves no hidden objects, and uses tactile instead of visual perceptual information. Children are shown a fake rock and asked what they think it is. After answering "a rock," they are allowed to feel that it is actually a sponge. Then the sponge is put away and the experimenter brings out a similar-looking real rock, saying, "Here is a real rock." Children are allowed to handle it briefly. They are then asked what another child who only looks at it will think it is. Children using PAR should reason that since the other child will not touch the object, he won't know that it is a real rock, and will get it wrong when asked whether it is a rock or a sponge. Therefore children using PAR should say that when the other child looks at the rock he will think that it is a sponge, while those using either RR or BR should once again pass the task.

The PAR hypothesis specifies that children only analyze the current situation to determine whether someone does or does not have perceptual contact with the object in question, and consequently whether he will be right or wrong about that object. For example, in the true and false belief transfer tasks, Maxi's return prompts children to see him as being in a *new* situation, and to use PAR about this new situation (i.e., "He doesn't see the object, so he doesn't know where it is, so he'll be wrong") without any reference to the prior situation in which Maxi acquired his true or false belief. Conversely, if Maxi did not leave, but simply placed his chocolate in the red cupboard and stayed in the kitchen, the simple disappearance of the chocolate into the cupboard would *not* constitute a new situation. The hypothesis specifies that when children do not decide that the situation has changed, they default to their initial conclusion that the protagonist will get it right.

While technically speaking it is true that at the moment an object disappears into a container, or a protagonist stops touching an object, he is out of visual or tactile perceptual contact with it, it would make no sense for children to re-apply PAR and conclude that he now doesn't know where or what it is and that he will be wrong about it. There would be no ecological validity to applying PAR on such a moment-by-moment basis. Furthermore, feedback from others' behavior would quickly extinguish any tendency to do so. The PAR hypothesis does not specify the cues children use to decide what constitutes a new situation, and thus what cues "trigger" the PAR rules, nor how these cues are learned (or whether they are innate). These are empirical questions.

Most importantly, the PAR hypothesis does specify that without a concept of mental representation there is no mental file, as it were, for children to put the information about what Maxi's perceptual access in the prior situation caused him to know, just as there is no mental file for them to put the information about what the visual appearance of a container or object causes someone to know about what is inside or what it really is. Consequently, the hypothesis should be easy to refute since it entails the unique prediction that children who use PAR will fail true belief tasks.[11] Specifically, it predicts a U-shaped developmental pattern of performance in true belief tasks: 3-year-olds should pass by using RR, 4- to 5-year-olds should fail by using PAR, and 6-year-olds should pass by using BR.

Fabricius et al (2010) tested this prediction with 108 children in three studies. In Study 1 they found the predicted pattern in the true belief transfer task and in the true belief contents task (see Figure 1). Note that in Figure 1 it appears as if 5½ -year-olds understand the location task, but not the contents task. Fabricius et al (2010) demonstrated in Study 2 that this difference is only apparent. The version of the true belief location task used in Study 1 inadvertently had twice as many references to the correct location than the incorrect one in the

[11] Two previous studies reported children's failure on true belief tasks (Friedman et al 2003, Ruffman 1996), although both sets of researchers assumed that children's difficulty with true belief was unique to the particular tasks used in their studies, and was not indicative of a general pattern of reasoning used across different tasks. See Fabricius et al (2010) for further discussion.

control questions, which apparently primed 5 ½ -year-olds (but not 4½ -year-olds) to choose the correct location. The researchers used the second version of the true belief location task in Study 2 which avoided that problem, and compared the results against a true belief contents task which was identical to the one used in Study 1. In Study 2, 5½ -year-olds scored the same on the two tasks (M = 65%), which falls favorably along the U-shaped function.

Fabricius et al (2010) also constructed and tested two methods for determining whether subjects are using RR, PAR, or BR. The first method involved giving each child both a false belief and a true belief task. Passing the false belief task and failing the true belief task indicates PAR. Passing the true belief task and failing the false belief task indicates RR. Passing both tasks indicates BR. The second method involved asking children for their justifications for their answers only on the true belief task (e.g., “Why will Elmo think that?”). Passing the task and giving a correct explanation (e.g. “Because it’s an M&M bag”) indicates BR. Passing the task but giving an incorrect justification (e.g. “Because it has M&M’s in it”) or no justification indicates RR. Failing the task regardless of justification given is sufficient to indicate PAR, because that is the only approach that gives the incorrect answer to true belief tasks.

The two methods yielded similar rates for each type of reasoning. For example, in Study 3, the first method (classifying children according to their performance on false and true belief tasks) indicated that 13% of 5½-year-olds used RR, 25% used PAR, and 58% used BR (only 4% failed both tasks). The second method (classifying children according to their justifications on the true belief task) obtained similar results: 17% of 5½-year-olds used RR, 29% used PAR, and

54% used BR. The average rates of BR across all three studies from Fabricius et al (2010) were 10% for 3½-year-olds, 29% for 4½-year-olds, 44% for 5½-year-olds, and 64% for 6 ½-year-olds.

The PAR hypothesis provides a consistent and principled explanation for the difficulty children who pass the standard, confounded false belief tasks have with 3-alternative versions of false belief tasks (Fabricius & Khalil, 2003) and with true belief tasks (Fabricius et al., 2010). It is principled because it provides clear criteria for distinguishing PAR from RR and BR. It is consistent because it provides the same explanation across all the task variations of the extant belief tasks. Tests of the hypothesis to date have also produced consistent results (except for Perner & Horn, 2003; but see Hedger, 2016, and Chapter 4, below). In contrast to the meta-analytic data showing that when they turn 5 years of age 75% of children pass the confounded false belief tasks (Wellman et al 2001), research using un-confounded tasks (Fabricius & Khalil 2003, Fabricius et al. 2010) has consistently found that only about 35% understand belief.

5. **The no belief task**

Most recently, Chen et al (2015) have garnered more evidence for the PAR hypothesis based on the No Belief Task.[12] This involved using cartoon pictures to tell a story to the subject. The teacher has brought a toy plane to class, which the students are all free to play with. She announces to the class that she is going to place the toy plane in one of two identical cupboards, but Lee (a student) has to leave to use the bathroom and misses where the teacher places the toy plane. When he returns, he wants to play with the toy plane. The test question is,

---

[12] Incidentally, this is also the first paper in support of the PAR hypothesis to come from outside the Fabricius lab. Fabricius and colleagues have also tested children on two different versions of the No Belief Task, and findings were consistent with the PAR hypothesis, but these data are yet to be published.

"Which cupboard will Lee open?"[13] Subjects were also asked about the likelihood that Lee would choose the other cupboard, and were asked for justifications for their response (as with Fabricius et al, 2010).

Of course, someone who understands the concept of belief (and is therefore using BR) should reason that Lee is ignorant of the toy's location, and so ought to be equally likely to choose either cupboard. Consistent with this, 59% of the 145 adults (across three experiments) chose the reality location (where the toy is currently located), which is not significantly different from chance. 97% of these adults said that Lee was just as likely to choose the other cupboard, and also gave appropriate justifications for their choice (e.g., "Perhaps toys were usually placed into this cupboard").

A subject using RR will choose the cupboard in which the toy plane is currently located, because that is where the toy really is. Consistent with this, of the children who failed the standard False Belief Task, 71% of them chose the reality location (where the toy is currently located). In contrast, a child using PAR should reason that Lee can't see the toy, so he doesn't know where it is, and consequently Lee will choose the incorrect location (i.e. the cupboard where the toy is not currently located). 80% of the 79 4-year-olds (across three experiments) did just this, and gave justifications consistent with the PAR hypothesis (e.g., "He didn't see where the toy was placed"). Only 25% of 4-year-olds said that Lee was likely to have chosen the reality cupboard instead. As the authors conclude—citing Fabricius & Khalil (2003), Fabricius et

[13] More strictly, in Experiment 1 the experimenters asked, "Which cupboard will Lee open first?" However, in Experiment 3 the 'first' was dropped (and a simpler story was used) in order to ensure that asking this question didn't confuse subjects. The results were consistent with what they found in the first experiment.

al (2010), and Hedger & Fabricius (2011)—"The results suggest that young children may use the 'perceptual access' approach in their theory of mind" (p. 79).

6. What we know about the false and true belief tasks (including an explication of the Rule A hypothesis)

The most plausible hypothesis concerning development of belief understanding needs to account for the data concerning both the True and False Belief Task in humans, summarized below in Table 1. (This table has been slightly over-simplified for ease of exposition. Note that here "False Belief Task" refers to the classic two-option versions described in Section 2 above.) Given that they are our closest living relatives, it would also be helpful to have a hypothesis which accounts for the data concerning the True and False Belief Task in chimpanzees, summarized on the first line of Table 1. The best theory might also help to explain some of the data concerning the autism spectrum disorder as well, such as that found on line 7 in Table 1.

A "?" in the table means that there are no published studies which I was able to find. The largest areas of understudied belief understanding involve the non-verbal true belief task in infants and 3-year-olds, and all non-verbal tasks in adults and school-age children. The study of the Non-Verbal True Belief Task in infants is contentious, because many researchers claim to have included such a task in their study, but Hedger & Fabricius (2011) explain that to be a true test of the PAR hypothesis, any such task requires a suitable cue for situation change, because children will only compute the PAR rules if they deem the present scenario to be a new situation (see Hedger & Fabricius 2011 for a review of some typical infant ToM studies).

Table 1.

	Non-Verbal False Belief Task[14]	**Non-Verbal True Belief Task**	**Verbal False Belief Task**	**Verbal True Belief Task**
Chimpanzees	Pass[15]	Fail[16]	N/A	N/A
Infants	Pass[17]	?	N/A	N/A
3-Year-Olds	Pass[18]	?	Fail[19]	Pass[20]
4- and 5-Year-Olds	?	?	Pass[21]	Fail[22]
6-Year-Olds	?	?	Pass[23]	Pass[24]
Neurotypical Adults	Pass[25]	?	(Presumably) Pass	(Presumably) Pass
Adults with High-Functioning Autism	Fail[26]	?	Pass[27]	?

[14] These non-verbal versions mostly follow the *Violation of Expectation (VOE) paradigm* (It is known that infants will look significantly longer at novel or unexpected stimuli than at stimuli which are familiar or expected; thus, if infants expect Maxi to go to one location, then they will look longer if Maxi goes to the other location) or the *Anticipatory Looking (AL) paradigm* (3-year-olds who say the incorrect answer will first look briefly at the false belief location). Some more recent studies use the *active helping paradigm* (subjects help agents by pointing to an object the agent is searching for). The chimpanzee studies involve a subordinate choosing whether or not to eat a prize based upon a dominant's perceptual access to the prize.

[15] Kaminski et al (2008), Hare et al (2001).

[16] Kaminski et al (2008).

[17] For reviews see Southgate (2014) and Carruthers (2013).

[18] Clements & Perner (1994), Garnham & Perner (2001), Garnham & Ruffman (2001), Ruffman et al (2001).

[19] Wellman et al (2001).

[20] Fabricius et al (2010).

[21] Wellman et al (2001).

[22] Fabricius et al (2010).

[23] Fabricius et al (2010).

[24] Fabricius et al (2010).

[25] Senju et al (2009).

[26] Senju et al (2009).

[27] Senju et al (2009). Children with autism typically fail these (Baron-Cohen et al 1985, Perner et al 1989).

According to the PAR hypothesis, a two-task battery (of a false and true belief task) can determine which type of reasoning a subject is using (Fabricius et al 2010, Hedger & Fabricius 2011). A subject using RR will predict that an agent will look for an object where it is actually located, and won't consider an agent's mental states or perceptual access. Thus, using RR should cause a subject to fail the false belief task and pass the true belief task. As explained above, a subject using PAR should pass the two-option false belief task (since the prediction is that agent will be unsuccessful due to a lack of perceptual access) and fail the true belief task containing an appropriate cue for situation change (once again because they predict that an agent will be unsuccessful due to a lack of perceptual access). Only a subject using BR, and making use of information concerning the agent's mental representations, will pass both tasks by correctly predicting the agent's actions.

Reality Reasoning (RR): **Fail the false belief task, Pass the true belief task**

Perceptual Access Reasoning (PAR): **Pass the false belief task, Fail the true belief task**

Belief Reasoning (BR): **Pass the false belief task, Pass the true belief task**

Turning to the verbal tasks on Table 1, we can see that 3-year-olds are using RR, 4- and 5-year-olds are using PAR, and 6-year-olds and neurotypical adults are using BR. These are the three developmental stages of the PAR hypothesis (Fabricius & Imbens-Bailey 2000, Fabricius & Khalil 2003, Fabricius et al 2010, Hedger & Fabricius 2011, Hedger 2016). We also need an explanation of the findings on non-verbal tasks, especially the now much-touted finding that infants pass non-verbal versions of the false belief task. Hedger and Fabricius (2011) conjecture that a developmental precursor of PAR, which they dub Rule A, may be at work in these infant studies. The two PAR rules (**see/not see → know/not know** and **know/not know → get it right/get it wrong**) are bridged by the concept KNOW, which children begin to acquire and link with perceptual access by about 3 ½ years of age (e.g., Pillow, 1989; Pratt & Bryant, 1990). Rule A is the condensed rule **see/not see → get it right/get it wrong**, which Hedger & Fabricius (2011) hypothesize children implicitly use before they acquire the concept KNOW.

The theory of Rule A follows directly from collapsing the two PAR rules used by preschoolers into a mechanism that could be used by nonverbal organisms such as chimps and human infants. The data summarized on the first line of Table 1 suggests that chimpanzees may use Rule A (see Hedger & Fabricius 2011 for more details), because Rule A (like PAR) should cause subjects to pass the false belief task but fail the true belief task. Hedger and Fabricius (2011) conjecture that Rule A may also be operative in human infants, and perhaps persists into adulthood.[28] More research into the anticipatory looking of children and adults would help to disconfirm or point in favor of the Rule A conjecture, as well as a non-verbal true belief task for

[28] In other words, Rule A may exist simultaneously in adult cognition along with BR, along the lines of a Dual Systems model such as Kahneman (2011). This could be tested by using the "eye gaze" methodology developed for children under 4 years of age (Clements & Perner, 1994; Garnham & Perner, 2001; Garnham & Ruffman, 2001; Ruffman, Garnham, Import, & Connelly, 2001), in which they have passed standard, confounded false belief tasks by showing unconscious anticipatory looking to the correct location. However, in order to know whether correct anticipatory looking in the false belief task indicates attribution of false beliefs or use of Rule A, the methodology needs to include a true belief task in which there is some interruption in the agent's connection to the situation that is comparable to what occurs in the false belief task. The previous eye gaze studies have not included such true belief tasks. If adults were found to continue to use Rule A implicitly it could explain a puzzling feature of theory of mind—that at times we appear able to make judgments about the mental states of others quickly, automatically, and effortlessly, while at other times the process is difficult and deliberative. There is evidence that BR is effortful and difficult for adults (Apperly et al, 2006; Keysar et al, 2003). Lin, Keysar and Epley (2010) found that higher working memory capacity can have a positive impact on adult performance in theory of mind tasks, while cognitive load impairs this ability. Perhaps this can be explained by BR and Rule A being different psychological mechanisms, the former explicit and effortful and the latter implicit, automatic and modular.

infants which includes a cue for situation change which is comparable to that found in the false belief task (such as having the agent temporarily leave and then return).

The popular *mindblindness* theory of autism spectrum disorder (ASD) is that part of the explanation of this disorder is an impaired ToM module (Baron-Cohen 1995).[29] The evidence for this includes the fact that children with ASD exhibit a lack of gaze-monitoring (Leekam et al 1993), failure to point in order to direct the attention of others (Baron-Cohen 1989a), delay on the unexpected transfer (Baron-Cohen et al 1985) and unexpected contents (Perner et al 1989) false belief tasks, delay on the appearance/reality task (Baron-Cohen 1989b), difficulty hiding objects (Baron-Cohen 1992), etc., despite the fact that children with intellectual disability (ID) and congenital blindness are able to successfully display most of these abilities (Baron-Cohen 1995).

Perhaps the innate mechanism missing in ASD is actually Rule A, as hypothesized by Hedger and Fabricius (2011). Higher functioning autistics may nonetheless be able to learn BR in order to eventually pass the False Belief Task. For instance, Senju et al (2009) found that adults with Asperger syndrome did not demonstrate anticipatory looking toward the correct location during the false belief task, although they verbally passed the task.[30] It should be noted that this kind of pattern has been found elsewhere in studies of ASD. For instance, Russo et al 2012 found that although multisensory processing times were not significantly different

---

[29] On modularity and psychology, see Fodor (1983) and Barrett & Kurzban (2006). On the hypothesis that ToM ability is the result of a module, see Baron-Cohen (1995), Carruthers (2013) and Leslie (1994). The details of Baron-Cohen's view are unimportant for our purposes, but it should perhaps be noted that his model includes four separate mindreading components, some of which (e.g. the Shared Attention Mechanism) are impaired in ASD while others (such as the Intentionality Detector) remain intact (Baron-Cohen 1995).

[30] Neurotypical adults *did* demonstrate anticipatory looking toward the correct location during the false belief task (which they also verbally passed), but a true belief task was not used in this study.

between persons with ASD and neurotypical subjects, brain activity (measured using EEG) *was* significantly different. As the authors note, this suggests that while on the surface persons with ASD and neurotypical persons appear to respond similarly to multisensory inputs, in actuality the brain mechanisms involved in this processing are entirely different (Russo et al 2012). Anecdotal evidence, such as Temple Grandin's well-known remark that she feels like an anthropologist on Mars when dealing with other people, also suggests that persons with ASD may compensate for certain impairments by using different methods; specifically, higher-level and less intuitive methods.

7. Troubles for the two dominant views in ToM

The current status of ToM literature in the psychology of cognitive development is largely a debate between two dominant views. On the one hand, the *Traditional View* is that children progress from RR directly to BR at about the age of four (e.g. Gopnik & Wellman, 1992; Perner, 1988, 1991; Saxe, Carey, & Kanwisher, 2004; Wellman et al, 2001). However, this theory faces difficulty explaining the finding that 4- and 5-year-olds fail the true belief task. If these children understand beliefs, then the true belief task should be passed at the same time as the false belief task. In fact, even though the task demands are made as commensurable as possible, intuitively the true belief task ought to be easier to pass than the false belief task.

On the other hand, the *Nativist View* is that ToM ability (including an understanding of belief) is innate, and furthermore the early onset of this capacity is evident in infancy. Although this view has been around since the 1980s (Leslie 1987, see also Fodor 1992), it has gained prominence in the past decade following Onishi & Baillargeon (2005)'s landmark study demonstrating that 15-month-olds can pass a non-verbal false belief (location) task. A flurry of studies has followed since Onishi and Baillargeon (2005) first cracked the dam (Buttelman et al 2009, He et al 2011, Scott & Baillargeon 2009; Song & Baillargeon 2008, Scott et al 2010, Surian et al 2007, Southgate et al 2010, etc). So many studies have been done at this point in time that it would be implausible to deny that infants are capable of passing a non-verbal false task (see Table 1). This also makes trouble for the Standard View.[31]

We have known for some time that 3-year-olds who gave the incorrect response on the false belief (location) task nonetheless often displayed anticipatory looking to the correct location (Clements & Perner 1994). According to the Nativist view the failure of 3-year-olds on false belief tasks is not a sign that they lack *competence* in using BR, but is simply a kind of *performance error* (Chomsky, 1965; Fodor, 1992) in which something interferes with this competence to cause them to fail false belief tasks. Leslie (1987; Roth & Leslie, 1998) hypothesized that true belief is the default attribution of our innate theory of mind mechanism, and that an inability to inhibit this default attribution in verbal tasks is what causes 3-year-olds to fail; in other words, failure on false belief tasks is due to misattribution of true belief. Leslie (2005) and others claim that infants do not exhibit the same difficulty inhibiting true belief

---

[31] See Perner & Ruffman (2005) for one attempt to salvage the Standard View in light of the infant studies; but see Chapter 5 below for an argument against this putative explanation.

Kevan Edwards and Bernard Kobes worry that later acquisition of belief understanding is not by itself evidence that belief understanding is not innate. I also accept this point. Although developmental psychologists that hold the Traditional View do not directly address this issue, I think they are assuming that (in the tradition of Piaget) when child development occurs in *stages,* the transitions between these stages involve a cognitive learning process, and therefore the understanding involved in the final stage is learned rather than innate. Also, Traditionalists tend to subscribe to the *theory-theory* view of mental state attribution (e.g. Gopnik & Wellman 1992, 1994), which is *prima facie* at odds with the view that mental state attribution is innate (once again because using a theory to attribute mental states is a deliberate cognitive process). (Granted, in Chapter 6 I will go against this tradition and suggest that the theories RR, PAR and BR may be innate, although there are clear stages of ToM development.)

attribution because the infant studies use eye gaze as a measure of understanding instead of verbal reports. Baillargeon et al (2010) also hypothesize that 3-year-olds fail verbal false belief tasks because of inhibition and selection failures, rather than a lack of belief understanding.

However, the true belief findings of Fabricus et al (2010) complicate this picture. Why would children fail true belief tasks if they have an innate theory of mind mechanism, especially one whose default attribution is that people have true beliefs? According to Roth and Leslie (1998), false belief tasks are more difficult than true belief tasks because of the need to inhibit the default true belief attribution, whereas this is not necessary in true belief tasks. Yet beginning at around age 4 children tend to pass false belief and fail true belief tasks. There is no obvious explanation for why 4- and 5-year-olds who are able to inhibit true beliefs in false belief tasks should inhibit true beliefs in true belief tasks as well. Thus the failure of inhibition explanation looks to be implausible, and so the finding that 4- and 5-year-olds fail the true belief task is a major problem for the Nativist view.

If it could be demonstrated that infants are able to pass a non-verbal true belief task, then Nativism might have a foothold. However, Hedger & Fabricius (2011) review the infant studies and explain that none of the true belief tasks used in those studies have a cue for situation change (such as a departure and return of the protagonist) which is comparable to that found in the false belief task. Hence it is possible that infants and toddlers are using a simplified, implicit version of Perceptual Access Reasoning (which Hedger and Fabricius 2011 dub “Rule A”) in order to pass the non-verbal versions of the false belief task. Hence, the PAR hypothesis of Fabricius and colleagues is best able to account for all of the ToM data.

Some may object that according to the PAR hypothesis, 3-year-olds are using Rule A implicitly (exhibited by their looking to the correct location on the false belief task) but using RR explicitly (exhibited by their incorrect verbal responses on the false belief task). Perhaps most surprising, that same implicit understanding will somehow rise to conscious reasoning in the next stage of development, as 4- and 5-year-olds begin to use PAR on verbal (false and true) belief tasks. This is indeed a puzzle to be explored, but its existence should not call into question the PAR hypothesis, because we see similar findings in other cognitive domains. Aside from the many dual-systems models of cognition in adults (e.g. Kahneman 2011, Kenrick & Griskevicius 2013), we also find infants demonstrating apparent understanding using looking time measures when they don't pass reaching procedures testing the same cognitive domain as toddlers.

For instance, infants pass VOE tests (see Footnote 11) of object solidity, but toddlers typically fail manual search versions of the exact same tasks, much like what we find with the false belief task. Spelke et al (1992) familiarized 4-month-olds with watching a ball hit the floor behind a screen. They then watched the ball being dropped behind the screen, after being shown that a physical barrier stood a foot above the floor. After dropping the screen, the infants looked significantly longer when the ball was shown on the floor compared to when the ball was shown on top of the physical barrier. This demonstrates that infants understand that solid objects cannot pass through other solid objects.

Hood et al (2000) used the same paradigm as the Spelke et al (1992) study, but this time they allowed toddlers to search in one of the two locations. Two-year-olds failed the task by searching for the ball on the floor instead of on top of the barrier. Susan Carey's (2009) explanation for this time lag sounds very similar to what Hedger & Fabricius (2011) hypothesize here in the case of the false belief task. She argues that in the looking time studies infants are able to pass by using "within module encapsulated representations;" their modular perceptual system represents objects as being solid and thus the infants are surprised when this expectation is violated (p. 113). However, when making a prediction two-year-olds are forced to use "explicit representations that are output of the perceptual device that creates representations of object-files" (Carey 2009, p. 113). In other words, actions such as searching for objects or predicting where objects will be found (or where others will look) forces 2- and 3-year-olds to use conscious reasoning processes which may rely on different representations than more modular, informationally encapsulated processes such as looking. At a later time (3-years-old according to Berthier et al, 2000) children are able to pass searching measures of object solidity, presumably because their conscious reasoning processes now involve the same representations as their more modular perceptual processes have used for at least 32 months. The same thing is happening with belief understanding, according to the PAR hypothesis.

To summarize, the PAR hypothesis can explain all of the findings summarized in Table 1, including some findings which pose difficulties for the two dominant views. The Traditional View, that children progress directly from RR to BR at about the age of four, is challenged by the finding that infants pass non-verbal 2-option false belief tasks. According to the PAR hypothesis, they do this by using Rule A (an implicit, modular mechanism which is a precursor of PAR). The Traditional View is also challenged by the finding that children ages four and five fail the true belief task and the 3-option false belief task. This is explained by the fact that there is an intermediate stage in between RR and BR in which children use PAR. The Nativist View, that BR is innate and the failure of 3-year-olds on the 2-option false belief task is a kind of performance error due to inhibition and selection failures, is also challenged by the finding that 4- and 5-year-olds fail the true belief task.[32]

According to the PAR hypothesis, infants pass VOE versions of the 2-option false belief task by using Rule A, which is presumably a modular or System 1 cognitive process. 3-year-olds fail the verbal 2-option false belief task, because they are using the conscious, System 2 reasoning process known as Reality Reasoning.[33] When children enter the next developmental stage around the age of four, and begin to use Perceptual Access Reasoning consciously and deliberately, they again pass the (verbal) 2-option false belief task. This same pattern of development which is part of the PAR hypothesis is also seen in studies of object solidity. 4-month-olds pass VOE measures of object solidity, which Carey (2009) hypothesizes is accomplished by using a modular cognitive-perceptual process. 2- and 3-year-olds fail searching versions of the same task, because they are using a different effortful and conscious process. By the age of four they pass the searching measures of object solidity, which Carey (2009) hypothesizes is because the new deliberate reasoning process now incorporates the same mental representations which were already being used inside the modular perception process.

8. Conclusion

[32] The PAR hypothesis also explains the performance of 4-year-olds on the no belief task (Chen et al 2015) as explained in Section 5 above, which presents yet another challenge for the Traditional and Nativists views.

[33] See Fodor 1983 for the modular/ conscious reasoning distinction, and Kahneman 2011 (and Chapter 4) for the System 1/ System 2 distinction. The vocabulary may vary somewhat, but I don't think the fact that there is some such distinction is at all contentious in psychology; on the contrary, it looks to be the standard view.

The PAR hypothesis best accounts for all of the data (the false and true belief task data summarized in Table 1, and the no belief task data discussed in Section 5). The studies demonstrating that infants are capable of passing the non-verbal false task is explained by the hypothesis that Rule A is an innate PAR mechanism. Further support for this conjecture is that our closest living relatives, chimpanzees, also appear to use Rule A, since they pass the false belief task but fail the true belief task. Still needed is a valid test of infants on the true belief task, as discussed in Section 5 above and Hedger & Fabricius (2011). The Rule A hypothesis would predict that infants fail the non-verbal true belief task, just as chimps do. Only the PAR hypothesis is able to explain why 4- and 5-year-olds fail the true belief task, even though they are able to pass the false belief task.

The philosophical upshot of all of this is that children do not have the concept of belief, and do not understand mental representation, before the age of 6. More evidence will be provided in the next chapter, where I examine children's understanding of perception. The point is made still more strongly in Chapter 5, when I turn to another current concern in cognitive science. Psychologists worry about finding an empirical method for determining whether chimps and infants are passing ToM tasks by using mentalist rules, which include a conception of mental representation, or whether they are passing them merely by using behavioral rules which require no understanding of the mind as such. Daniel Povinelli (2001; Povinelli & Vonk, 2003, 2004; Penn & Povinelli, 2007) has made the argument that there is in fact no empirical method for deciding which of these two explanations is true, and Josef Perner (2010, forthcoming) has recently applied the argument to the case of human subjects. In Chapter 5, I show that PAR allows us to solve this problem. The solution makes clear that children under the age of six don’t understand mental representation, and therefore lack the adult concepts of perception and belief.

[Some of this chapter is a revision of material which originally appeared in Hedger & Fabricius (2011) and Hedger (2016). Portions of this chapter were also presented at the Midwest Empirical and Theoretical Association for Philosophical Research Meeting at the University of Illinois, Urbana-Champaign, on the 26th of April, 2012. Thanks to participants who gave comments and criticisms; in particular Bill Fabricius, Renee Baillargeon, Denise Cummins and Robert Cummins. Thanks also to Bill Fabricius, Bob Van Gulick, Kevan Edwards, Bernie Kobes and Kim Frost for comments on earlier drafts of this chapter.]

References

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17,* 841-844.

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

Baron-Cohen, S. (1989a). Perceptual role-taking and protodeclaritive pointing in autism. *British Journal of Developmental Psychology, 7,* 113-127.

Baron-Cohen, S. (1989b). Are autistic children behaviorists? An examination of their mental-physical and appearance-reality distinctions. *Journal of Autism and Developmental Disorders, 19,* 579-600.

Baron-Cohen, S. (1992). Out of sight or out of mind: Another look at deception in autism. *Journal of Child Psychology and Psychiatry, 33,* 1141-1155.

Baron-Cohen, S. (1995). Mindblindness: An essay on autism and theory of mind. Cambridge, MA: MIT Press.

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

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review, 113,* 628-647.

Berthier, N. E., DeBlois, S., Poirier, C. R., Novak, M. A., & Clifton, R. K. (2000). Where's the ball? Two- and Three-year-olds reasons about unseen events. *Developmental Psychology, 36,* 394-401.

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

Carey, S. (2009). *The origins of concepts.* Oxford, England: Oxford University Press.

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

Chen, Y., Su, Y., & Wang, Y. (2015). Young children use the "ignorance = getting it wrong" rule when predicting behavior. *Cognitive Development, 35,* 79-91.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: The MIT Press.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*, 67-90.

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

Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy, 60,* 685-700.

Davidson, D. (1975). *Inquiries into truth and interpretation.* Oxford, England: Oxford University Press.

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences, 1*, 568-570.

Dennett, D. C. (1981). True believers: The intentional strategy and why it works. In A. F. Heath (Ed.), *Scientific explanation: Papers based on Herbert Spencer lectures given in the University of Oxford.* Oxford: Clarendon Press, pp. 150-167.

Fabricius, W. V., Boyer, T., Weimer, A.A., & Carroll, K. (2010). True or False: Do five-year-olds understand belief? *Developmental Psychology*, *46*, 1402-1416.

Fabricius, W.V., & Imbens-Bailey, A.L. (2000). False beliefs about false beliefs. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning about the mind* (pp. 267-280). Hove, England: Psychology Press.

Fabricius, W. V., & Khalil, S. L. (2003). False beliefs or false positives? Limits on children's understanding of mental representation. *Journal of Cognition and Development, 4*, 239-262.

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

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind.* Cambridge, MA: The MIT Press.

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition, 44,* 283-296.

Friedman, O., Griffen, R., Brownell, H., & Winner, E. (2003). Problems with the seeing = knowing rule. *Developmental Science, 6*, 505–513.

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

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.

Gonzales, C., & Fabricius, W. V. (2013). Four-year-olds' performance on three-alternative false belief tasks: Evidence for perceptual access reasoning. Poster presented at Cognitive Development Society VIII Biennial Meeting, Memphis, 2013.

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7,* 145-171.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour, 61*, 139–151.

Harman, G. (1978) Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences 1*, 576–77.

He, Z., Bolz, M., & Baillargeon, R. (2011). False-belief understanding in 2.5-year-olds: Evidence from violation-of expectation change-of-location and unexpected-contents tasks. *Developmental Science 14*, 292–305.

Hedger, J. A. (2016). Perceptual access reasoning: Developmental stage or System 1 heuristic? *Phenomenology and the Cognitive Sciences, 15,* 207-226.

Hedger, J. A., & Fabricius, W. V. (2011). True belief belies false belief: Recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development." *The Review of Philosophy and Psychology, 2*, 429-447.

Hogrefe, G..J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attributing epistemic states. *Child Development, 57*, 567 – 582.

Hood, B., Carey, S., & Prasada, S. (2000). Predicting the outcomes of physical events: Two-year-olds fail to reveal knowledge of solidity and support. *Child Development, 71,* 1540-1554.

Horgan, T., & Woodward, J. (1985). Folk psychology is here to stay. *Philosophical Review, 94,* 197-225.

Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin Books.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition, 109*, 224 – 234.

Kenrick, D. T., & Griskevicius, V. (2013). *The rational animal: How evolution made us smarter than we think.* New York: Basic Books.

Keysar, B., Lin, S. & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25-41.

Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., and Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology, 15,* 77-95.

Leslie, A. M. (1987). Pretense and representation: The origins of 'theory of mind.' *Psychological Review, 94,* 412-426.

Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In *Mapping the mind: Domain specificity in Cognition and Culture,* Ed. L. Hirschfeld and S. Gelman. Cambridge, England: Cambridge University Press.

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

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy, 50,* 249-258.

Lin, S., Keysar, B., & Epley, N. (2010). Refelexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology, 46,* 551-556.

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

Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind." *Philosophical Transactions of the Royal Society B—Biological Sciences, 362,* 731-744.

Perner, J. (1988). Developing semantics for theories of mind: From propositional attitudes to mental representation. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 141-172). New York: Cambridge University Press.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: The MIT Press.

Perner, J. (2010). Who took the 'cog' out of cog*nitive* science? Mentalism in an era of non-cognitivism. In P. A. Frensch & R. Schwarzer (Eds.), *Cognition and neuropsychology: International perspectives on psychological science, Volume 1,* pp. 241-261. ??,??: Psychology Press.

Perner, J. (forthcoming). Theory of mind—an unintelligent design: From behavior to teleology and perspective. In A. M. Leslie & T. C. German (Eds.), *Handbook of 'Theory of Mind'.* Piscataway, NJ: Lawrence Erlbaum Associates.

Perner, J., Frith, U., Leslie, A., & Leekam, S. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development, 60,* 689-700.

Perner, J. & Horn, R. (2003). Knowledge or false negatives: Do children of 4 to 5 years simulate belief with "Not knowing = getting it wrong?" *Journal of Cognition and Development, 4,* 263-273.

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

Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology, 47,* 116–129.

Poulin-Dubois, D., Sodian, B., Metz, U., Tilden, J., & Schoeppner, B. (2007). Out of sight is not out of mind: Developmental changes in infants' understanding of visual perception during the second year. *Journal of Cognition and Development, 8*, 401-425.

Povinelli, D.J. (2001). On the possibilities of detecting intentions prior to understanding them. In B. Malle, D. Baldwin, & L. Moses (eds.), *Intentions and Intentionality: Foundations of Social Cognition.* Pp. 225-248. Cambridge, MA: MIT Press.

Povinelli, D.J. & Vonk, J. (2003). Chimpanzee minds: Suspiciously human? *Trends in Cognitive Science, 7*, 157-160.

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

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development, 61,* 973–982.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515-526.

Pylyshyn, Z. (1978). When is attribution of belief justified? *Behavioral and Brain Sciences, 1*, 592-593.

Ramsey, W., Stich, S. P., & Garon, J. (1990). Connectionism, eliminativism, and the future of folk psychology. *Philosophical Perspectives, 4,* 499-533.

Roth, D., & Leslie, A. M. (1998). Solving belief problems: Toward a task analysis. *Cognition, 66,* 1-31.

Ruffman, T. (1996). Do children understand the mind by means of simulation or a theory? Evidence from their understanding of inference. *Mind and Language, 11*, 388 -414.

Ruffman, T., Garnham, W., Import, A., & Connelly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology, 80*, 201-224.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515-526.

Russo, N., Mottron, L., Burack, J. A., & Jemel, B. (2012). Parameters of semantic multisensory integration depend on timing and modality order among people on the autism spectrum: Evidence from event-related potentials. *Neuropsychologia, 50,* 2131-2141.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55,* 87-124.

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

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

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in asperger syndrome. *Science, 325,* 883-885.

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

Southgate, V. (2014). Early manifestations of mind reading. In S. Baron-Cohen, H. Tager-Flusberg & M. Lombardo (Eds.), *Understanding Other Minds*, 3rd Edition. Oxford University Press.

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

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

Spelke, E. S., Bilinger, K., Macomber, J., & Jacobsen, K. (1992). Origins of knowledge. *Psychological Review, 99,* 605-632.

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

Vonk, J. & Povinelli, D.J. (2006). Similarity and difference in the conceptual systems of primates: The unobservability hypothesis. In E. Wasserman and T. Zentall (Eds.) *Comparative Cognition: Experimental Explorations of Animal Intelligence.* pp. 363-387. Oxford University Press.

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.

Wellman, H. M., Lane, J. D., LaBounty, J., & Olson, S. L. (2011). Observant, Nonaggressive Temperament Predicts Theory of Mind Development. *Developmental Science*, *14*(2), 319–326.

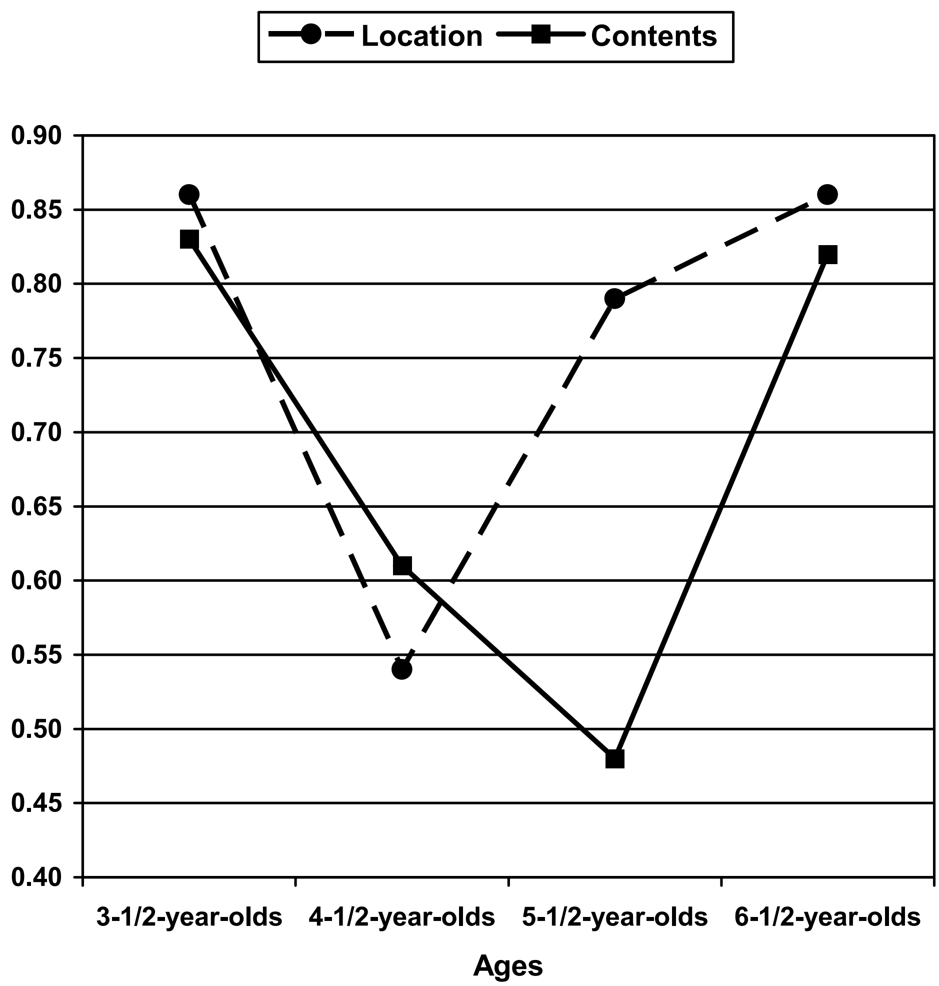
Williamson, T. (1995). Is knowing a state of mind? *Mind, 104*, 533-565.

Williamson, T. (2000). Knowledge and its limits. Oxford, UK: Oxford University Press.

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.

Figure 1. Children's Performance on True Belief Tasks

(From Fabricius, W. V., Boyer, T., Weimer, A.A., & Carroll, K. (2010). True or false: Do five-year-olds understand belief? *Developmental Psychology, 46,* 1402 – 1416, American Psychological Association, reprinted with permission.)



CHAPTER THREE

Seeing:

The Perceptual Access Reasoning hypothesis and young children's understanding of perception

1. Introduction

After examining the child's understanding of belief through the lens of the Perceptual Access Reasoning (PAR) hypothesis, which I have argued best explains all of the existing data (Chapter 2), it becomes obvious that children do not acquire a concept of mental representation until around the age of six (this idea will be developed further in the chapters which follow). Thus we can begin our examination of the child's understanding of perception by sketching what such an understanding would be like if one lacked a conception of mental representation, and then looking at the picture which emerges from the large amount of data which developmental psychologists have gathered on children's understanding of perception over the last few decades and determine whether that picture is consistent with our sketch. In this chapter I will first present that sketch briefly here in the introduction, before examining the empirical evidence in the sections which follow. In particular, we will look at studies pertaining to John Flavell's highly influential model of the four levels of perception understanding in Section 2, studies pertaining to the child's ability to distinguish reality from the way things appear to be in Section 3, and finally some closing remarks in Section 4.

If we take away the concept of mental representation, then what kind of an idea of perception might we be left with? First, this notion of perception is likely to be more directly causal, like one pool ball smacking into another. One is reminded of Aristotle's metaphor of an object making an impression on the mind like a signet ring makes an impression on a piece of wax (*De Anima* 424a 18-20). Without the representational relationship between the mind and the world, we are left with a much more simplistic, unmediated and direct relationship. One key idea tied to the adult understanding of perception that will be obviously missing here is the distinction between veridical and mistaken perceptions. Without representation there cannot be misrepresentation. Just as it doesn't make sense to talk of the wax impression misrepresenting the seal which was pressed into it, neither should it make sense to the child to talk about a person seeing something that isn't there or experiencing a visual illusion. Thus the work by Flavell and others on the Appearance/ Reality Distinction will be crucial when deciding the plausibility of the PAR hypothesis.

There shouldn't be much more to the PAR child's conception of visual perception than mere visual contact. Thus Gopnik and Wellman (1994) write that "very young children seem to treat desire and perception as simple causal links between the mind and the world… Given that an object is within a viewer's line of sight, the viewer will see it." This seems to be corroborated by anecdotal evidence. I remember when my son was 4 for example and he couldn't seem to understand how I could not see one of his toys when I was facing in the direction of it, even though it was just one item amidst a chaotic assortment of small toys piled on the floor. And there is the well-known phenomenon that children will "hide" things in plain view, as when my son used to put a penny into his right palm right in front of me and then ask me to guess which hand was holding the coin. Perner (1991) relates a similar story of how Heinz Wimmer's 3-year-old son would "hide" in the pantry right in front of Wimmer when playing hide-and-go-seek (p. 153). Although children understand that a viewer cannot see an object which is occluded from her line of sight by about 2 and ½ years of age (Flavell et al 1978), the boys in these anecdotes (and many similar ones) don't appear to understand that their fathers retain a visual representation which persists after the environment has been altered to obstruct their line of sight. This reasoning can be seen as consistent with the first of the two PAR rules (see previous chapter)—Daddy doesn't see the penny (in the sense that it is not in his direct view; Daddy lacks unmediated perceptual access to the penny) so he won't know which hand it's in. Or, Daddy doesn't see me (again in this simplistic sense) so he won't know I'm in the pantry.[34]

Second, this simpler notion of perception is going to be much more closely tied to the immediate environment. With no mental file with which such subjects can carry perceptions which persist through new situations (see Chapter 2), perceptions should change at the exact same rate as the surroundings of the perceiver change for the PAR child.[35] For instance, consider the true belief task (Chapter 2). Maxi watches Mom move his chocolate from the blue cupboard to the yellow cupboard before going outside outside to play. There is no subsequent movement of the chocolate. However, for the child using PAR, when Maxi returns he does not

---

[34] For the PAR rules to be a full explanation of this type of phenomenon, however, it must be that the occluding of the coin and the hiding in the pantry constitute cues for situation change. Since these sorts of behaviors are more typical of 3-year-olds than 4- and 5-year-olds, who are more likely to be in the PAR stage of cognitive development, perhaps this is indicative of a transition from Reality Reasoning (RR) to PAR—more complex than RR but without the more sophisticated notion of a "new situation" that we see 4- and 5-year-olds use in the true belief task. (See Chapter 2 for more about cues for situation change in PAR.)

[35] More precisely, perceptions change whenever both the environment changes *and* there is a sufficient cue for the subject (see Note 1 above and Chapter 2).

perceive the chocolate. Maxi has no representation of the chocolate in the yellow cupboard because PAR does not involve representation. Maxi's visual perceptions—what Maxi sees—are directly related to his immediate environment or situation. And in the current situation Maxi has no visual perception of the chocolate. He had a visual perception of the chocolate in the yellow cupboard, but that was connected to the old situation, and that perception vanished as soon as the situation changed.

Of course, even the adult says of this scene that Maxi cannot currently see the chocolate because it is not in plain view; it is inside the cupboard. However, for the adult using Belief Reasoning (BR), Maxi just saw the chocolate a moment ago when he watched Mom put it in the yellow cupboard. Maxi retains the *mental representation* of the chocolate in the yellow cupboard. However, for the PAR child this "just saw the chocolate a moment ago" also vanished when the situation changed. Maxi of course has no beliefs about the chocolate for a subject using PAR, because what is a belief if not a representation? But even for the PAR child, "just saw the chocolate a moment ago" is relevant to their reasoning about Maxi, *as long as the environment does not change* for Maxi. If Maxi watches Mom move the chocolate from the blue cupboard to the yellow cupboard and does not go outside to play, then the child using PAR will predict that Maxi will search the yellow cupboard if he looks for his chocolate. He just saw the chocolate being put into the yellow cupboard.[36] However, this "just saw the chocolate" disappears when the environment changes to a new situation.

[36] Indeed, recall that on my view this is why most infant theory of mind studies fail to demonstrate use of Rule A (a precursor of PAR); their true belief conditions lack a sufficient cue for situation change (Hedger & Fabricius 2011; see also Chapter 2).

Children in the PAR stage should nonetheless be able to understand occlusion, distance effects, etc., and even the sort of perspective-taking involved in realizing that I can see an object that you cannot see—after all, I can imagine what might happen if something impacted my body without considering mental states at all. Consider the findings of Yaniv and Shatz (1988), who argue that failures of young children on perception tasks such as the ones I report in Sections 2 and 3 is due to "their difficulty in accessing their knowledge in particular circumstances. This proposal contrasts with a common view in the field [and argued in this chapter] that young children are incapable of making their inferences about the mental states of others" (p. 95). They support their proposal with a study demonstrating that 3- 4- and 5-year-olds understand (to some degree) that certain factors can affect perceptual access. For example many of them answered correctly that distance between a subject and object can prevent the subject from smelling or touching the object but not seeing or hearing the same object; or that occlusion separating an otherwise nearby subject and object can prevent seeing and touching the object but perhaps not hearing the object.

However, for the child in the PAR stage of theory of mind development, determining perceptual access is a rather straightforward process which doesn't require information about the person's mental states. Predicting that another person won't be able to see an occluded object is akin to predicting that I won't be able to toss a ball into your hands if you're standing behind an object which completely covers your body. It's a purely physical causal process in this sense, which only requires understanding the environment and drawing imaginary lines between objects and a person (or a person's eyes in the case of sight, a person's hands in the case of touch, etc.). Therefore, the findings of Yaniv and Shatz (1988), and many other researchers making similar claims, is in fact quite compatible with the inability of young children to make inferences about mental states.

2. Flavell's Four Levels of Perception Understanding

Flavell's (1974) classic developmental model of perception understanding distinguishes four different levels. *Level 0* involves understanding object permanence, or the expectation that an object can be seen again once the child moves around an occluding object. In other words, it is the understanding that objects persist even when they are no longer perceived. In the language of Piaget (1954, Piaget & Inhelder 1956), the *Level 0* child has a pragmatic and practical sensory-motor understanding of objects in space, and at some knowledge-in-action level "knows" that the same object appears differently depending upon the child's own perspective and distance from that object. However, the child lacks any symbolic-representational understanding of visual perspective (and therefore can't really think about them or have conscious expectations about them).[37] The *Level 0* child is also egocentric in Piaget's sense, in that she doesn't consider or recognize that there are visual points of view other than her own.

Piaget (1954) assumed that children reach *Level 0* (which he called the acquisition of the "object concept") at around 12 months, since younger infants typically make the *A-Not-B* error. That is, after an object is hidden in one location (A) and repeatedly found there, the object is hidden in a second location (B) and the infant is allowed to search. Before about 12 months of

[37] Perhaps we could put this into the more modern language of Carey (2009) and Fodor (1983) by explaining that the *Level 0* child has only modular representations of visual perspective which can't be accessed by conscious reasoning systems.

age, infants typically continue to search for the object in the first location (i.e. searching in A instead of in B). However, later researchers assume that these sorts of perseveration errors are performance failures that don't demonstrate true lack of competence (see Chomsky 1965 for the performance/competence distinction). That is, they are due to lack of inhibition rather than genuinely not understanding that the object persists in location B. More recent studies find evidence of object permanence understanding as early as 3 ½ months of age (Baillargeon & DeVos 1991). For our purposes, we should note that something more than *Level 0* understanding is required even for PAR, since the rules take as input whether another person has perceptual (e.g. visual) access to a given object. *Level 0* children are not capable of considering the visual point of view of others.

Level 1 involves understanding that other people may or may not be able to see the same objects that the subject sees. However, this is still not an understanding of mental representation, because the *Level 1* child does not consider that another person has a unique point of view or visual perspective which may differ from that of the child. The *level 1* child does not understand that the same object can appear differently to different subjects, but only reasons about objects and whether or not they fall within a person's line of sight. For instance, one task measuring *Level 1* perspective taking, the *Picture Task* (Flavell 1974, Masangkay et al 1974), involves seating the child at a table across from the experimenter. The child is shown a card which has two different pictures on each side, for instance a dog on one side and a cat on the other. After familiarization with the card and the pictures, the experimenter places the card in between the child and himself, and asks "What do you see? What do I see?"[38] Masangkay et al (1974) found that about 65% of 2.5 -year-olds (Experiment 1), and all 3-year-olds (Experiments 1 and 2) were able to pass this task.

A similar task conducted by Flavell and colleagues (Task IIIE in Flavell et al 1968) was found to be more difficult. The experimenter first familiarizes the child with a card which is identical on both sides; it contains pictures of an airplane on top, a teddy bear in the middle, and a clown on the bottom. The experimenter then demonstrates how a piece of cardboard can be used to occlude one or more of the pictures on one side. Placing the card with the pictures in between the subject and the experimenter, and placing the cardboard over pictures on the experimenter's side (the cardboard is larger than the picture board so that the child is able to infer where it is placed), the child is asked, "Can you tell me what pictures I see on my side?" The cardboard is then lowered and the test question is asked a second time. Only 2 of 10 3-year-olds answered both questions correctly, but most of the 4- and 5-year-olds did, and the 6-year-olds performed perfectly (i.e. none of them answered a test question incorrectly).

Level 1 understanding is clearly required for PAR. For example, in the belief tasks, the subject must consider what Maxi sees or does not see in order to compute the first PAR rule (see Chapter 2). Of course, *Level 1* by itself is not sufficient for PAR, since 3-year-olds who pass the Picture Task nonetheless fail the standard 2-location false belief task. Although necessary for PAR, *Level 1* does not require an understanding of mental representation. In other words,

---

[38] Alternately, the subject is asked, "Do you see the cat, or do you see the dog? Do I see the cat, or do I see the dog?" No significant differences in success have been found between this version of the test question and the version given in the text above.

Level 1 understanding is (of course) not sufficient for BR either. For example, a *Level 1* subject may be able to tell that a given person cannot see an object because that person's line of sight is obstructed. However, she also expects someone who has line of sight to an object to see that object, not understanding that a person can look in the direction of something but still not see it, for example because it is camouflaged in some way. The findings on *Level 1* tasks are thus completely commensurable with the PAR hypothesis, and my claim that an understanding of mental representation is not acquired until the BR stage, or around 6 years of age (see Chapters 2, 5 and 6).

Level 2 involves understanding that the same object can appear differently to different people. Of course, the *Level 1* child recognizes that the dog in the Picture task is seen by the subject but not by the experimenter. However the *Level 2* child understands that an object which is visible to both the subject and an experimenter can look different because of each person's different perspective. Unlike the previous two levels, this would seem to involve an understanding of mental representation, since the *Level 2* child must understand that visual representations of the same object can differ across subjects. That is, *Level 2* involves realizing that different people can see the same object *in different ways* at one and the same time—that different people can have different mental (visual) representations of the same object.

As tests of *Level 2* vary quite a bit in complexity and task demands, so do the ages when subjects typically pass them. Piaget used the *Three Mountains task* (Piaget & Inhelder 1956), which involved familiarizing the child with a three-dimensional scale model of three mountains of different size in a nonlinear arrangement. After a doll was seated so that it was facing the same model from a different visual perspective (e.g. across from and facing the child), the subject is asked to imagine what the mountains look like from the doll's perspective. The child was asked to select from among several alternatives the photograph which best reproduced the doll's visual perspective. Piaget and Inhelder (1956) found that children as old as 10 had difficulty with this task (although some children as young as 7 were successful), and Laurendeau and Pinard (1970) found that only 28% of 12-year-olds passed (the oldest age group they tested). Flavell et al (1968) complicated the task further. In their version, a child was shown a three-dimensional nonlinear arrangement of three cylinders of different sizes, but the cylinders were also painted red on one side (i.e. for half of their circumference) and white on the other. The subject was then asked to reconstruct a given view of the cylinders using a duplicate set. Less than half of the 16-year-olds tested passed this task.

Borke (1975) modified Piaget and Inhelder's (1965) Three Mountains Task in a couple of ways in order to test 3- and 4-year-old subjects. She used a toy car with the Sesame Street character Grover inside, which the experimenter "drove" around one of three different three-dimensional model displays on square tables. Display 1 was a lake with a toy sailboat, a miniature cow and horse, and a model house. Display 2 was Piaget and Inhelder's Three Mountains model. Display 3 was a farm scene with a number of miniature animals and people, a barn, a windmill, and a lake with ducks in it. The experimenter would then stop when the car with Grover was facing a different side of the display than the child, for a total of three times per display for each subject. The subject had an exact duplicate of the display on a table which she could turn. When the car with Grover was "stopped" by the experimenter pretending to drive it, the child was asked to move the turntable display until she was looking at it the same way Grover is. Using the Three Mountains display (Display 2), the 3-year-olds were correct 42% of the time and the 4-year-olds were correct 67% of the time, already an improvement over the success rates of Piaget and Inhelder (1956). However, using Display 3 (the farm scene with multiple objects), the 3-year-olds gave correct responses over 79% of the time and the 4-year-olds were correct 93% of the time.

The simplest *Level 2* task is probably the *Turtle task* (Masangkay et al 1974). This test involves placing a picture (of a turtle in this case) flat on the table when the child and experimenter are seated across from each other, so that one person sees the turtle with its feet on the top and its shell on the bottom, while the other person sees the reverse. The test questions are “Do you see the turtle right side up, or upside down? Do I see the turtle right side up, or upside down?” Masangkay, Flavell and their colleagues (Masangkay et al 1974, Experiment 2) found that 25% of 3-year-olds, 50% of 3.5-year-olds, and 100% of children aged 4 and older passed this task. This would mean that children in the PAR stage of ToM development should successfully pass the Turtle task. However, 3-year-olds consistently fail this task, and Flavell and colleagues (Flavell et al 1981) demonstrated that increased familiarization, retesting, and training all fail to improve their low success rates.

Level 2 is intended to be the dividing line which requires an understanding of mental representation. *Level 1* only involves understanding how to draw a line of sight from a person’s eyes to a given object, so the subject need not consider the subjective visual experience of the other person, or their unique visual point of view. However, when we move to *Level 2*, the subject is now putatively considering the visual perspective—the mental (in this case visual) representation—of the other person. As Flavell (1974) explains, “it must be admitted that the main focus of [the *Level 1* child’s] attention in the person-object relation is still on the object rather than the person. In all probability, his representations decrease in salience and articulatedness as they go from the more external to the more internal aspects of that relation; that is, from object seen, to [the other person’s] visible looking gestures (movements and positioning of head and eyes, ‘intent’ expression, etc.), to [the other person’s] covert visual experience. In contrast, *Level 2* knowledge seems unquestionably to be knowledge of precisely these latter, purely internal and phenomenological aspects” (p. 97).

However, for at least some of the *Level 2* tasks, I think we can question whether understanding of mental representation is truly required. For example, in Borke’s (1975) experiment, the subject may be simply reasoning about which objects in the three-dimensional display are closest to Grover’s line of sight, considering the character’s head orientation and proximity but not considering how (or even that) Grover represents the scene. The subject could then pass the task by simply turning the table until that object which appears closest to Grover is then closest to the child. This could explain why children found manipulating the three-dimensional model of the three mountains (Borke 1975) much easier than they found selecting the other character’s perspective from among a group of photographs (Piaget & Inhelder 1956).

In fact, as other researchers have noted (e.g. Zatchik 1990), there is an explicit similarity between photographs and mental representations in that photographs are representations as well. If subjects are indeed considering the visual representation of the other person in order to pass these tasks, then it would seem to be a straightforward and simple process to match their simulated visual representation with the corresponding photograph. However, 4- and 5-year-olds find this task to be very difficult. Fishbein et al (1972) also found that manipulating physical objects was much easier for young children than selecting photographs. They used a revolving tray containing toys which differed depending upon which side of them a person was viewing, for example a toy soldier holding a candy cane in its left arm and saluting with its right arm. In the *Turning task*, the experimenter sat directly across from the child and asked the subject to turn the tray so that the experimenter could see a specified view of the toy; e.g. "Show me the side of the soldier holding the candy cane." In the *Pointing task*, the experimenter sat in one of four positions, so he was facing either the front, left side, back, or right side of the toy. The child was then asked to select from a series of 8 photographs (showing the toy from 1 of 8 different angles) the one which best matched what the experimenter sees. The subject was given corrective feedback (e.g. shown which photograph was correct) when she responded incorrectly, and rewarded with an M&M candy for correct responses. The 4-year-olds gave correct responses on the *Turning task* over 90% of the time, but their percentages of correct responses dropped to only 48% on the *Pointing task*, and dropped further to 34% when 3 toys were used instead of only 1.[39]

---

[39] Even when the experimenters reorganized the data in order to count the selection of an adjacent angle as correct, so that when the experimenter views the toy from the front e.g. a child who selects the photograph depicting the toy straight on, or the photograph depicting the toy from a 45 degree angle, or the photograph depicting the toy from a 315 degree angle (so that 3 of the 8 photographs would be counted as a correct response), the 4-year-olds gave correct responses only 50% of the time for the 1 toy display and 39% of the time for the 3 toy display (Fishbein et al 1972, p. 25).

One way to account for this dramatic difference between success rates depending upon whether subjects had to turn the tray or select a photograph, is that in the *Turning task* a child is simply moving the tray so that an object is visible to either the subject or the experimenter. For instance, when the experimenter instructs the child to "Show me the side of the soldier holding the candy cane," the child simply reasons about the experimenter's line of sight and moves the tray so that the candy cane is visible to the experimenter but not the child, in much the same way that a subject in the *Picture task* (Masangkay et al 1974) might flip the card around so that the cat is visible to the experimenter but not the subject. In this way the *Turning task* can be seen to be nearly identical in inference demands to the *Picture task*, so that what is actually being tested is *Level 1* (and not *Level 2*) understanding.

In contrast, the *Pointing task* explicitly asks the child to imagine the visual perspective of the experimenter. The child must predict what the toy looks like to the other person, i.e. the experimenter's mental (visual) representation, in order to pass this task. If the child could indeed do so, it should then be a relatively simple matter to match that simulated visual representation with the corresponding photograph. However, in reality the subjects found this to be much more challenging. Thus, the *Turning task* tests *Level 1* understanding but the *Pointing task* tests *Level 2* understanding of perception. Reinterpreting the results in this way, we can see that the 4-year-olds (presumably in the PAR stage of ToM development) demonstrated *Level 1* understanding but not *Level 2* understanding. Hence they do not truly demonstrate understanding of mental representation, when the data are interpreted correctly.

If all of this is correct, then the turning tasks of Borke (1975) should also be interpreted as testing *Level 1* and not *Level 2*.[40]

Similarly, in the *Turtle task* (Masangkay et al 1974), the subject may be reasoning about which part of the turtle (the feet or the shell) is closest to the experimenter's direct line of sight, rather than considering the visual perspective of the experimenter. In their *Fishes task* (Experiment 2) meant to test *Level 1* perception understanding, three "Mr. Potato Fish" toys are arranged on a display so that one is always closer than the other two to a stationary observer. Subjects were asked to select from three identical toys the one which corresponds to the fish that the child or the experimenter "sees best." According to the authors, "In both the pilot work and the present study, [subjects] took the one they 'see best' to mean the one closest to and facing them, as we had intended they should" (p. 361). In this *Level 1* task, all three fish are visible but subjects are asked to select which fish the experimenter sees best. In a similar fashion, subjects in the *Turtle task* may be simply considering whether the experimenter sees the turtle's feet or the turtle's shell best. Hence the task may not actually require an understanding of mental representation.

Although I am confident in my evaluation of the turning tasks of Borke (1975) and Fishbein et al (1972) as tests of mere *Level 1* understanding, and hence versions of the tasks asking subjects to choose from a selection of photographs to be more accurately assessing their understanding of mental representation, there are at least two points against our interpreting away the *Turtle task* findings (Masangkay et al 1974) in a similar fashion. First, subjects found

[40] Of course, neither Borke (1975) nor Fishbein et al (1972) explicitly reference the developmental model of Flavell under consideration here, which was published two years after the latter paper.

the *Fishes task* to be easier than the *Turtle task*, suggesting that the two tasks are not testing the same cognitive process.[41]

Second, their Experiment 3 contained two tasks involving toy witch heads which were explicitly designed to test this interpretation (see p. 364). From the front the witch's long protruding nose and open mouth were clearly visible, while from the back the witch's cone hat was visible but not her nose and face. In one task (called *Witch-split*) the toy witch was cut in half, and the front half containing the nose was pasted onto one side of the card and the other half containing the cone hat was pasted onto the other side. The test questions were, "Do you see the witch's nose or the witch's hat? Do I see the witch's nose or the witch's hat?"[42] The other task (*Witches*) involved placing a witch between the child and the experimenter, who were seated across from each other, so that it faced one of them and had the back of its head facing the other person. The subject was then asked to choose from a display of three toy witches, one facing towards the child, one facing away from the child, and one with its side profile facing the child. The test questions were, "Which one of these witches looks exactly like what you see? Which one of these witches looks exactly like what I see?"

Although *prima facie* similar in task demands, *Witch-split* explicitly asks subjects to think about the witch in terms of its parts, and so is reasoning about what objects a person sees (the nose or the hat), while *Witches* explicitly asks about how another person sees the same object as the subject (what it "looks like" to them), and so is reasoning about another's subjective

[41] Whereas 50% of 3.5-year-olds gave correct responses in the *Turtle task*, all of them passed the *Fishes task*. And whereas 25% of 3-year-olds gave correct responses in the *Turtle task*, two thirds of them passed the *Fishes task.*
[42] Note the similarity to their aforementioned *Picture task* utilizing the board with a picture of a dog one side and a picture of a cat on the other.

visual point of view. All of the 3.5-year-olds passed *Witch-split*, but none of them passed *Witches*. If children reasoned about which part or object (the nose or the hat) the experimenter saw best in the *Witches task*, then it should be as easy as the *Witch-split task*, but clearly it's not. This is rather compelling evidence that subjects are not passing the *Turtle task* by considering whether the feet or the shell or closest to the experimenter's line of sight. (The Turtle task also explicitly asks whether the experimenter *sees* the turtle right side up or upside down.) Although it's possible that the same story we gave for the turning tasks (Borke 1975, Fishbein et al 1972) also applies to the *Turtle task*, it starts to look rather implausible in light of this experiment.

However, when we consider the data on *Level 2* tasks as a whole, all of it is consistent with the view that children in the PAR stage of ToM development (i.e. roughly 4- and 5-year-olds) lack an understanding of mental representation, with the sole possible exception of the *Turtle task* (Masangkay et al 1974). Furthermore, the wide variation of ages at which children pass the various Level 2 tasks—from age 4 for the *Turtle task* (Masangkay et al) to older than 7 (Piaget & Inhelder 1956) or 12 (Laurendeau & Pinard 1970) for the *Three Mountains task*, to older than 16 for the three cylinders task (Flavell et al 1968)—suggests that Flavell's (1974) initial hunch about *Level 2* understanding is likely to be true: "It seems reasonable to suppose that, here as elsewhere in the domain of cognitive development, there are numerous intermediate and transitional sublevels" (p. 97). The understanding that the same thing can appear differently to different people, that people can and do have distinct visual representations of the same object, and predicting what that other person's visual representation is like, of course admits of *degree*. So in some ways it's not at all surprising that we see such wide variation in the ages at which subjects typically pass *Level 2* tasks.

If Flavell is right that the *Level 2* stage itself admits of sub-stages of development, then perhaps PAR children are at the beginning end of that spectrum, having not yet acquired the concept of mental representation but slowly developing it over the next couple of years. This would make sense given that PAR is a transitional stage of ToM development between RR and BR. Also, the PAR rules presented last chapter don't have an obvious application to any of the *Level 2* tasks, so that they cannot test whether children are using PAR or not in the way that our false belief task-true belief task battery can (see Chapter 2). Hence my conclusion is that the findings of *Level 1* and *Level 2* understanding in children is completely consistent with the PAR hypothesis, and the view that an understanding of mental representation—which separates BR children from those in the PAR stage—is not acquired until around six years of age.

Finally, the *Level 3* child not only recognizes that the same object can appear differently to different people, and can not only predict how an arrangement appears from another's perspective, but is also able to predict features of the other person's retinal image. For instance, a Level 3 subject is able to reproduce the projective (versus real) sizes of objects from another's perspective, noting for example that although one object is really bigger than a second object, it will appear smaller to another person because it is farther away from him. There are few, if any, developmental studies which test this level of perception understanding. Since *Level 3* is of little interest for present purposes, I mention it here for completeness' sake but have nothing further to say about it.

3. PAR and the Appearance/Reality Distinction

After working on this model of perception understanding, Flavell and his colleagues spent years exploring when children learn that reality can be different than it appears, which they dubbed *the appearance/ reality distinction* (Flavell et al 1983). For our purposes, the ability to understand that such a distinction can be made would seem to be an essential aspect of having the concept of mental representation. The fact that we represent the world introduces a key mediation between ourselves and the world. Without mental representation, there is only us and the world, and we have a simple relationship of either having (perceptual) access or lacking access to the world. It is more like a physical relationship; for instance I either have access to my apartment or I don't (e.g. when it's locked and I don't have the key, or when I'm in a different location).[43] Presumably, the PAR child's understanding of perception is a similarly simple model. Either a person has visual access to a scene or object, or she lacks access. Just as with me and my apartment, as long as a person is in the vicinity and there is no obstruction such as her eyes being closed or her line of sight being occluded by another object, then she has visual access to the object in question.

However, the adult understanding is more complex than this due to the concept of mental representation.[44] Now there is not simply us and the world, but there is us, the world,

---

[43] Of course, I'm not claiming that young children are behaviorists or lack any conception of mind whatsoever. Even this simple notion of *seeing* is not completely devoid of mental concepts. The key thing children in the PAR stage lack is an understanding of *mental representation*.

[44] It should perhaps be noted here that even most adults have a naïve understanding of perception and memory which greatly simplifies the role that our brain has in creating our perceptions and memories (we spend a lot of time talking about this in my Critical Thinking class). Understanding perception, like understanding a Platonic dialogue or anything else, comes in degrees. It is not a binary operation of either understanding or not understanding, like a light switch being either on or off. However, lacking any conception of mental representation is more serious than not having a full understanding of perception. This idea is so fundamental to understanding

and our representations of that world. With this less direct, more mediated access to the world, representations of the world can be accurate or inaccurate. We can correctly represent some aspect of the world as it really is, or we can incorrectly represent some aspect of the world. Now beliefs can be true or false, and perceptions can be veridical or non-veridical. Now we understand that things can appear different than they really are. In the words of Flavell et al (1983), "[W]e have acquired the metacognitive knowledge that appearance—reality differences are always among life's possibilities" (p. 96). Hence examining when children acquire the appearance/ reality distinction, besides being interesting in its own right, is crucial to evaluating the plausibility of the PAR Hypothesis. More specifically, prior to looking at the empirical data, we would predict that children should not have a full understanding of the appearance/ reality distinction in the PAR stage of development (i.e. roughly ages 4 and 5).

3.2 Appearance/ Reality Tasks

Probably the most well-known task Flavell and his colleagues (1983) used to test understanding of the appearance/ reality distinction involved showing children a sponge which looks like a rock, and then allowing them to handle it. After the child is familiarized with the fake rock, it is placed on the table. Children are then asked, "When you look at this with your eyes right now, what does it look like?" and "What is this really, really? Is this really, really a rock or really, really a sponge?" (p. 102). Hereafter, I will refer to this as the *Standard Appearance/ Reality Task.* They found that before four years of age children tend to give the same answer for both questions (i.e. they say that it really is a sponge, but also that it looks like

---

perception that lacking it causes a difference in *kind* between the young child and the adult rather than a mere difference of *degree*.

a sponge), presumably because they don’t realize that things can be different than they appear to be.

Flavell and colleagues used a number of variations of this paradigm to test children’s understanding of the appearance/ reality distinction, including using objects such as rubber pencils and candles that look like apples, or changing the appearance of ordinary objects. For example, they would place a white index card behind a blue plastic filter, or put a straight object into a clear glass container filled with water, or view objects through a magnifying glass. Their creativity led to a great number of variations, including e.g. the experimenter pouring coffee into a mug from a watering can, with a flower placed in front of the mug, so that from a distance it looked as though the experimenter was watering a plant but in actuality he was pouring coffee into a cup. Experiment 3 of Flavell et al (1983) included a total of 21 different Appearance/ Reality (A/R) tasks asking children about the identity of imitation objects, the size, color and shape of various objects, how far away objects were, objects hidden behind larger objects which occluded them, and actions such as the one described with coffee in a watering can.

3.3 Results

Unsurprisingly, children in what Hedger and Fabricius (2011) dub the RR stage of development tend to fail these tasks, and children in what they dub the BR stage tend to pass. As Flavell (1986) notes, “Only a few three-year-olds get them right consistently, whereas almost all six- to seven-year-olds do” (p. 419). Unfortunately, the performance of children in the PAR stage—i.e. 4- and 5-year-olds generally—is a bit more complicated. Initially, Flavell et al (1983) ran three experiments. In Experiment 1, performance of 4 ½ -year-olds (mean age 4 years, 8 months) varied from 65% to 95% giving correct answers to both the appearance and reality question.[45] In Experiment 2 , success rates of 4-year-olds varied from 19% to 43%, and 5-year-olds were correct 55% to 73% of the time. Finally, in Experiment 3, 4 ½ -year-olds (mean age 4 years, 5 months) answered both questions correctly about half of the time.

As we saw previously with the Level 2 tasks, performance on A/R tasks varies considerably for children ages 4-7. Thus on a more difficult variation of the standard A/R task (Flavell et al 1986), 6 ½-year-olds answered correctly 45% of the time, and answered correctly while giving appropriate verbal justifications for their answers which included reasoning about appearance and reality only 12% of the time (Study 5).[46] Flavell et al (1986) concluded that 6 ½-year-olds

> ...seem to find it particularly difficult to talk about appearances, realities, and appearance –reality relations... even briefly and minimally... They also tend not to mention them when asked to administer the very sorts of standard AR tasks they find so easy, as subjects, to solve—even after the experimenter has explained and repeatedly demonstrated the administration procedure... [W]e believe these difficulties in verbal labelling and nonverbal identification at least partly reflect genuine conceptual difficulties. Many subjects of this age simply seem unable to think about notions of "looks like," "really and truly," "looks different from the way it really and truly is," and so on... (pp. 58-59)

[45] 5 ½ -year-olds (mean age 5 years, 10 months) were almost perfect; notice however, that these children are on average almost 6 years old, and so most of them are probably in the BR stage.

[46] An attempt to improve performance of 6 ½ -year-olds (in Study 6) through a pre-training procedure and by repeating "more clear and explicit task instructions" before every task (for a total of 23 times) was unsuccessful (p. 45).

Thus we have another puzzle in the form of apparently conflicting data about the success rates of children ages 4-7 on A/R tasks, much as we had in the previous section with Level 2 tasks. However, before we turn to attempting to solve this puzzle, let's first take a look at children's *failures* on A/R tasks, and what they mean from the perspective of the PAR theory we are developing.

3.4 Errors on A/R Tasks: Phenomenism and Intellectual Realism

As mentioned previously, children who fail the standard A/R tasks tend to give the same answers to both the appearance and the reality questions, rather than getting both questions wrong. For example, 3-year-olds tend to say that the rock-sponge looks like a sponge and really is a sponge, that the stick viewed partially submerged in water or through a prism looks bent and really is bent, and so on. More specifically, Flavell et al (1983) note that when asked about an object's *properties*—such as the shape of the stick or the color of an index card or the size of an object viewed through a magnifying glass—3-year-olds tend to answer both questions based upon the current appearance of the object in question. Thus they say for example that the index card looks blue and really is blue when seen through a piece of blue-colored plastic. This is known in the literature, following the terminology of Piaget and Inhelder (1956), as *Phenomenism*. In contrast, when asked about the *identity* of an object, 3-year-olds tend to answer both questions based upon the current reality of the object in question. Thus they say that the rubber pencil both looks rubber and really is rubber, rather than a pencil. This is known in the literature, again following the terminology of Piaget and Inhelder (1956), as *Intellectual Realism*.

Since 3-year-olds tend to answer both the questions the same, then that means they get one of the test questions correct. When asked about the properties of an object, they tend to answer the appearance question correctly but the reality question incorrectly, and when questioned about the identity of an object, they tend to get the appearance question wrong but the reality question correct. Thus Flavell and colleagues (1983, Flavell 1986) seem to interpret these results as partial understanding of the appearance/ reality distinction. They are partially correct, after all. However, the fact that children don't give random answers to the test questions doesn't imply partial understanding of the distinction at all.

If we keep in mind that what these children lack is a conception of mental representation, then of course there is no distinction between what something looks like and how it really is. These amount to the same thing. The only possible response for such a child is to answer both questions the same (e.g. it looks like a sponge and really is a sponge). To answer both questions incorrectly would mean that children can make a distinction between how something appears and the way it really is, but simply that they're confused about which representation is the appearance and which is the reality. When we think about it carefully, we see that in fact we should predict that a child who truly doesn't understand mental representation and therefore can't make the reality/ appearance distinction will give the same answer to both test questions.

Flavell et al (1983) found that when asked about properties, young children who fail the standard A/R task tend to exhibit phenomenism and that when asked about identities the same children tend to give intellectual realist responses. This may make perfect sense when we consider the way our brains understand the world. We seem to be hard-wired to believe that identities of things remain constant, but that the properties of those objects may change over time, and indeed usually do change (given sufficient time). Consider Descartes's example of a piece of wax. Initially, Descartes observes that the piece of wax is hard and cold when he touches it, makes a sharp tapping sound when he raps it on the table, has a certain shape that he can see, and—being recently taken from the beehive—still retains an odor of the flowers and a faint taste of honey. However, as he starts to melt the wax by placing it near the fire, all of these properties change. It is now warm and soft to the touch and makes hardly any sound when struck. The shape and color and even the former size of the piece of wax now change, so that it's visual appearance is drastically different from a moment before. What was left of the smell and taste also now evaporate as it melts. Descartes observes that all of the properties of the piece of wax have changed right before our senses, but we continue to identify it as the same piece of wax. The properties have changed, but the identity remains the same (Descartes 1641/1985, pp. 20-21).

We see myriad examples of the phenomenon in our everyday world, and young children also recognize that it's the same ball even though the shape changes after it's been deflated, that it's the same maple tree whose leaves change from green in the summer to amber and orange in the autumn, and who sheds those same leaves in the winter, the baby becomes the girl becomes the woman becomes the old woman but remains the same person, and on and on. (Incidentally it appears to be this world where the properties of things are constantly in flux which so engaged the Ancient Greek philosophers, and which is the root of many metaphysical identity puzzles in philosophy.)

Hence we see that it would be perfectly natural for the child lacking a concept of mental representation to apply these beliefs about ever-changing properties to the Appearance/Reality tasks. Since there is no distinction between reality and the way our minds represent reality, then there is only one thing, viz. reality. And as we have already seen, subjects who fail the A/R tasks consistently give the same answer to the appearance and the reality question. The object or the property of the object looks the same as it really is, since without mental representation there is no difference between the thing and the way it appears.

What's more, as we see from our reasoning above, identity remains constant but an object's properties change. Therefore it would make sense for the child to make the phenomenism error when asked about properties and the intellectual realism error when asked about identities. The stick looks bent and it really is bent (phenomenism) for the child who lacks a concept of mental representation. A moment ago it looked straight and really was straight. This is actually perfectly reasonable given the current conceptual repertoire of the child. There is no distinction between the property of an object and the way we represent that property, there is only reality. And in reality, the properties of things change. The handlebar of the bicycle becomes bent, the ball goes flat, the shape of the wax changes when it gets hot. The shapes of objects are changing all of the time. So in the A/R tasks, for the RR or PAR child the index card is blue when seen behind the blue filter and it is white when the filter is removed. The color of a sunset changes before our very eyes as we watch it; why should index cards be any different? The adult explanation, that the filter makes the white card *look* blue temporarily, is not available to these subjects.

In contrast, the identities of objects remain constant although their properties are continually in flux. The same piece of wax remains through its numerous property changes, and the same toy soldier remains after it is bent. Thus it would make sense that the same sponge remains regardless of its properties. Therefore the young child says that it not only really is a sponge but it also looks like a sponge right now (intellectual realism). Colors and shapes and sizes may change, but identities do not change. If the object is a sponge, then that's reality and it is always a sponge. Even adults would agree with the reasoning so far. Since there is no distinction between the sponge and how it appears or is visually represented, then it must also look like a sponge. Hence the phenomenism and intellectual realism errors are perfectly reasonable and predictable, given that we keep in mind that the child lacks the concept of mental representation, and that our brains view the world as one filled with objects whose identities remain constant but whose properties may change.

There is also some evidence that young children may view properties as more malleable than adults do, so that from their point of view properties are constantly changing from situation to situation. One example is Piaget's well known *conservation task* (Piaget & Inhelder 1974). This can take many different forms, but a typical task is the following. First, children are shown two identical glass containers, each holding an equal amount of liquid, and are asked if the two amounts of liquid are the same. Of course, most children successfully answer "Yes." While the child observes, the experimenter takes the liquid contents of one of the glasses and pours it into a taller and narrower glass. Piaget, and many other researchers who successfully replicated his findings, found that many children aged 2 to 6 years will now say that the two amounts of liquid are *different*, because the thin glass contains *more* liquid than the original glass contained.

Of course adults believe that the amount of liquid in a glass can change, for instance if someone spills some of a glass of bourbon or takes a drink out of it. However, an adult who witnessed the bourbon being poured from one glass to another would never think that the amount of bourbon has somehow changed simply as a result of the act of pouring it into a new glass, regardless of the phenomenal appearance of either glass of bourbon. Hence, children at this stage of development seem to think that properties of objects can change from situation to situation, and those different situations may be as simple as being in different containers. (This was further corroborated by the fact that children made similar mistakes in other cases, such as when a spherical ball of clay is rolled into a snake-like shape, etc.)

Another example is that infants may believe that the steepness of a ledge changes from situation to situation. At about 1 year of age many infants transition from crawling to walking. Kretch and Adolph (2013) found that although 12-month-olds who crawl over a shallow cliff (e.g. 9 cm) will not attempt to crawl over a 90 cm cliff (0% attempted this), 63% of 12-month-olds who are walking will attempt to walk over the 90 cm cliff (i.e. approximately 3 feet). Of course, there are many possible interpretations of this finding, but one possibility is that infants think the same cliff is less deep when walking than it is when crawling. Finally, it appears that young children may believe that the stick actually does bend when submerged in water. Given what they believe about the world and the mind, this conclusion actually appears to be the optimally rational one. Children lacking the concept of mental representation may begin with false premises, but their reasoning is quite logical.

3.5 The PAR Stage of Development and Success on A/R Property Tasks

Now that we understand that for young children the properties of objects can change from situation to situation, we may have a clue how to solve our puzzle about the performance of 4- and 5-year-olds on A/R tasks (most of whom are using PAR). Perhaps 4- and 5-year-olds who pass the standard A/R tasks, by answering e.g. that the stick looks bent but is really and truly straight, mean by "really and truly straight" only that the stick will be straight again once it is lifted out of the water, but that it really is bent right now. In other words, whereas the BR child or the adult mean that the stick simultaneously appears bent but is actually straight, because they can hold two conflicting representations of the stick in their minds at one time, the PAR child reasons only about how the stick is in reality, but believes that it changes from straight to bent back to straight again when submerged and then lifted out of the water. Thus the conflicting representations are not concurrent but rather chronological—the stick was straight, now it is bent, and it will be straight again. In fact, there is a significant amount of evidence in favor of this latter interpretation of what 4- and 5-year-olds are saying.

First, in many cases when 4- and 5-year-olds are asked basic questions about how things are, without explicit reference to appearance or reality, they often answer based upon the appearance of the object (i.e. phenomenism) rather than the way the objects really are. Russell and Haworth (1988) had two subjects witness an illusory display and then asked them questions about it. For instance, in one task two children were first shown an array of white objects on a table, and then instructed to wear glasses with red Perspex lenses. The test question was, "What color are these things on the table?" 60% of the 70 5-year-olds tested (mean age 5 years, 2 months) responded with "red" rather than "white," in contrast to the 70 7-year-olds (mean age 7 years, 2 months) who only responded "red" 25% of the time.[47] Thus when asked about the color of the objects, 60% of the 5-year-olds answered based upon phenomenal appearance rather than reality, whereas only 25% of the 7-year-olds did this.

Braine and Shanks also did a number of experiments which are designed to test Piaget's idea of *conservation*[48] but are very interesting for our purposes here (1965a, 1965b). The *size* experiment is as follows. Two simple figures were cut out of colored construction paper—for instance a square, star, ball, or house shape. For each pair of cutouts, there would be an identical shape (e.g. square), but each of the shapes would be different color (e.g. one red and one blue) and they would also be of different sizes (roughly 1" and 1.5"). A viewing apparatus was constructed with two separate compartments, each with its own viewing screen in which either an ordinary pane of glass, or a magnifying lens which doubled the apparent size of objects, can be placed. (The glass pane and magnifying lens were removable so that the experimenter could surreptitiously switch them in order to be sure that trials were counterbalanced.) For each trial, the subject first viewed the two cutouts (e.g. a 1" red square and a 1.5" blue square) outside the viewing apparatus, and then watched as the experimenter

---

[47] Attempts to use social pressure to push 5-year-olds towards giving a realist answer had no effect. In order to control for the hypothesis that the phenomenism may be a result of the test context, Russell & Haworth (1988) had experimenters ask for a single response from the two subjects, and then gave one of them glasses with red Perspex lenses and the other glasses with clear lenses, so that answering "clear" would be the best way to resolve differences of opinion. However, even on these trials 5-year-olds said "red" 60% of the time.

[48] See section 3.4, above.

placed both cutouts into the compartments. The experimenter always placed them so that the 1.5" blue square was placed behind the simple pane of glass and the 1" red square was placed behind the magnifying lens, so that it now looked larger (specifically 2") than the 1.5" blue square. Subjects were then asked, "Which *looks* bigger?" and, "Which is *really, really* bigger?" (emphasis in the original). (The questions were of course counterbalanced across trials.)

The *shape* experiment involved first showing the subject two rods of different colors, one of which is bent and the other of which is straight. These were then immersed in water so that the refraction made the bent rod appear straight and the straight rod appear bent. Subjects were then asked, "Which *looks* bent?" or, "Which *looks* straight?" (counterbalanced across trials); and, "Which is *really, really* bent?" or, "Which is *really, really* straight?" (the adjective matching that of the first test question).

Braine and Shanks found that, across a series of four experiments using these two paradigms (1965a, 1965b), roughly 35% of the 4-year-olds and 65% of the 5-year-olds were able to answer most of the test questions correctly (specifically, at least 9 out of 10 test questions).[49] However, 4- and 5-year-olds often gave answers based upon appearance rather than reality when asked a neutral question which didn't include "looks" or "really." In Experiment IV of Braine & Shanks 1965b, subjects were given eight trials of the *size* experiment, above. After viewing the cutouts outside of the viewing apparatus and then inside the viewing compartments, subjects were asked, "Which is the big one?" Following each trial, the experimenters used a feedback procedure in which they were told either "Yes, that's right; this

[49] See Figure 3 on p. 204 of Braine & Shanks 1965b.

is the big one…See? [taking the cutouts back out of the viewing apparatus],” or “No, that’s wrong; this is the big one…See?” Subjects earned two poker chips for correct answers or had one taken away for incorrect responses, and the poker chips could be traded in for M&M candies following the experiment (which subjects knew from a pre-training procedure).

In the first trial, 100% of the 4- and 5-year-olds answered incorrectly. That is, they all said that the square which is smaller in reality, but looks bigger behind the magnifying lens, was the big one.[50] After being told they were wrong in the feedback procedure following the task, and with the motivation of earning poker chips which they could use to “purchase” candy, performance increased on the second trial of the same task to about 35% correct for 4-year-olds and 40% correct for 5-year-olds. These are the same results Russell and Haworth (1988) got when they asked children similar questions about the color of objects, above. Even after being given feedback and correction on their answers, nearly two thirds of the 4- and 5-year-olds in Braine and Shanks’ (1965a) Study IV nonetheless continued to respond with the cutout which *looks* bigger behind the magnifying lens when asked, “Which is bigger?” One explanation of this finding (and the similar finding from Russell & Haworth 1988) is that the subjects believe that the properties of the objects (e.g. the size and color) literally do change when viewed through a distorting lens or colored eyeglasses.

From the adult perspective, the cutout remains the same size but we visually represent the object as being larger than it actually is when viewed through a magnifying glass. However,

[50] An interesting contrast is that 15% of the 3-year-olds answered correctly on the first trial (see Braine & Shanks 1965a, p. 234). However, it’s not known whether this difference is statistically significant given the number of subjects (7 three-year-olds, 11 four-year-olds, and 12 five-year-olds).

to the 4- and 5-year-olds, who are using PAR and have no concept of *mental representation,* this explanation makes no sense. From their perspective, there is only the object in reality and a subject's perceptual access (or lack thereof) to the object in question. For them, there is only the cutout which they can see, whose properties can change over time. When viewed through the magnifying glass, therefore, the cutout actually *is* bigger than it was a moment ago. There is no "looks bigger" from the standpoint of PAR; there is only "bigger." So of course when asked, "Which is bigger?" the 4- and 5-year-olds answer that the one which appears bigger is the larger one. They believe it is the larger one, when it is behind the magnifying lens. When it is pulled back out of the viewing apparatus, then it is again the smaller object.

Since there is no "looks bigger" for the children using PAR, but only "is bigger," then how do they interpret the questions from the Standard A/R tasks, such as "Does the stick look bent or does it look straight [when viewed through a prism]?" For the PAR child, there is only reality (the "is"), and the subject's perceptual access to that reality (or their lack of perceptual access to that reality). There is no mental representation, including visual, so there is no "looks straight" for them. Given their limited conceptual repertoire, they are forced to interpret the question "Does the stick look bent or does it look straight?" as meaning "Is the stick bent or is it straight?" For the PAR child, properties are probably situation-dependent in the way we noted above. Hence the stick is in reality bent when viewed through a prism or partially submerged in water. Thus 4- and 5-year-olds answer that the stick looks bent in the standard A/R task. This is the correct answer from the adult perspective, but means something completely different in the mind of the child using PAR. If they lack an understanding of representation and therefore don't distinguish appearance from reality, then children using PAR may interpret "looks" and "is" to mean the same thing.

Similarly, as we hypothesized at the outset of this section, children using PAR may also interpret the standard A/R task *reality question,* "Is the stick really and truly bent or is it really and truly straight?" as a question about how the stick is at other times, i.e. how the stick was previously and how it will be in the future (before and after being partially submerged in water). The adult of course understands that "really and truly" is meant to contrast with how the stick appears, i.e. how it is visually represented. For the adult, the stick simultaneously is visually represented as bent but is straight in objective reality. For the child using PAR, there is no visual representation, but the stick actually is bent in reality when partially submerged.

Thus they answer to the first question that the stick 'looks' (meaning 'is') bent. Since the second question is seeking a contrast, the only contrast for the child using PAR is how the stick is in reality in different situations. For them, properties are situation-dependent. So "really and truly" must mean, for the child using PAR, how the stick is in other situations, i.e. how it is when not partially submerged in water. Thus they answer that the stick is "really and truly straight," but again they mean something different than the adult does by this locution, given that their conceptualizations differ.

In fact, this is the same theory of five-year-olds put forth by Flavell and colleagues (1989), independently of the PAR hypothesis. In a pilot study, they showed 5-year-olds two different displays—one with illusory color and one with illusory shape. However, they asked different reality questions for each display (counterbalanced across subjects). For one display, they asked the same reality question as in the standard A/R task: "*Really and truly*, is this stick *really and truly* bent or *really and truly* straight [emphasis in the original]?" However, for the other display, they asked, "How about right now? *Right now* is this stick *really and truly* bent or *really and truly* straight?" They found that 5-year-olds "gave approximately twice as many incorrect answers to the 'right now' reality questions as they did to the standard reality questions" (p. 514). Hence they seemed to interpret the standard A/R reality question in the way we suggested, as meaning "How is the stick in other situations?" However, when prompted with the locution "right now," 5-year-olds were twice as likely to say "bent" when the stick was partially submerged, because for them the stick is in reality bent 'right now' in the present situation.

Flavell et al (1989) then tested 24 adult college students and 24 5-year-olds on the appearance/reality questions, measuring how often subjects spontaneously used "looks" when answering the appearance question and "really and truly" when answering reality question. They found that 65% of the adults and 61% of the children spontaneously used "really and truly" in answering the reality question, but that whereas 85% of the adults used "looks" when answering the appearance question, only 29% of the 5-year-olds did. Flavell et al (1989, pp. 519-520) note that

> The 29% is consistent with findings from our pilot study and from previous research that children of this age tend not to employ "looks" or "looks like" spontaneously when referring to illusory displays (Flavell et al, 1986; Russell & Haworth, 1988). It is also consistent with our hypothesis that they tend not to think of appearances as appearances, but rather as present realities. Because they tend to equate "looks" and

> "is," they find it natural to use the more common term "is" when describing appearances.

Of course, if they lack an understanding of mental representation, then these 5-year-olds may lack the cognitive ability to distinguish appearances from reality, and so they are forced to equate the two.

Furthermore, the locution "right now" included in the appearance question may cue the children to think of the present situation. Thus, they answer that the stick looks (is) bent, because that's how it is in the present situation. If anything, the "looks" may simply be a reminder that the subjects (from their point of view) have perceptual access to the stick in the current situation, because they are looking directly at it. However, just as children of this same age appear to reason that quantities can change from situation to situation in Piagetian conservation tasks, they may also reason that the stick changes from straight to bent and back again from situation to situation. The remaining salient difference between the two questions (once "looks" and "is" are equated) then becomes "really and truly," which may cue 5-year-olds to think of situations other than the present one—that is, past (and perhaps future) situations where the stick is straight.

Flavell et al (1989) also asked subjects another question when the stick was partially submerged, which is not typically found in standard A/R tasks. One of the test questions was, "Is the stick a different shape than it was before?" Only 20% of the college students in the study answered that the stick was a different shape than it was a moment ago (p. 518). Presumably most of the adult subjects (if not all of them) reason that although the stick appears bent right now, the actual shape of the stick is nonetheless still straight. However, roughly three quarters of the 5-year-olds answered “Yes,” that the stick is a different shape now than it was a moment ago when it was not partially submerged. In fact, *70% of the 5-year-olds who passed the standard A/R task answered that the stick is a different shape than it was before.*

This is evidence that although they are answering the standard A/R questions correctly in the sense that they are giving the same answers as adults, 5-year-olds (who are typically using PAR) do not interpret the questions the same way that adults do. For the child using PAR, the stick is straight before submerged, bent while partially submerged, and straight when taken back out of the water. There is no mental representation, visual or otherwise, and so there is no distinction between the way an object appears and the way it really is. Therefore the data gathered from A/R tasks is perfectly consistent with the PAR hypothesis which we set out last chapter. What’s more, our theory can explain the findings of Flavell et al (1989), Russell and Haworth (1988), and Braine and Shanks (1965a, 1965b), which are problematic for the view that 4- and 5-year-olds understand mental representation and have acquired a theory of mind. The PAR hypothesis also supplies the mechanism by which 4- and 5-year-olds misinterpret (from the adult perspective) the standard A/R test questions in the way in which Flavell and his colleagues (1989) hypothesized, just as the discovery of DNA revealed the mechanism by which Darwin’s hypothesis of heritable traits works in reality.[51]

---

[51] Flavell et al (1989) actually continued to think that children in this stage do understand mental representation to some extent. Thus they “do not believe that they [i.e. 5-year-olds] would be actually incapable of construing the stimulus as being simultaneously, at this moment, characterisable in two different, contradictory-seeming ways” (p. 513). In this way, the view presented here contrasts sharply with that of Flavell et al (1989), since without the concept of mental representation children in the PAR stage are literally incapable of thinking of the stick as both bent and straight at one and the same time. The only way adults do that is by distinguishing objective reality (straight stick) from visual (mis)representations (looks bent).

3.5 The PAR Stage of Development and Success on A/R Identity Tasks[52]

Since for the PAR child properties are continually in flux and relative to the environmental situation, while identities remain constant, they must interpret the test questions in the A/R identity tasks differently than they do in the A/R property task. Recall that 4- and 5-year-olds tend to pass these tasks as well. In the Standard A/R task which uses the fake rock, children tend to answer correctly by responding with “a rock” when asked, “When you look at this with your eyes right now, what does it look like? Does it look like a rock or look like a sponge?” They also give the correct response of “a sponge” when asked, “What is this really, really? Is this really, really a rock or really, really a sponge?”

The correct answer to the reality question should come as no surprise to us, because as we saw earlier children interpret “really, really” simply as across contexts; i.e. at different times and in different environmental situations. So the sponge is always a sponge and remains so across contexts. Just as with Descartes’ piece of wax, it’s the same sponge whether the subject has perceptual access to it or not, and whatever its visual appearance is.

However, how are they to interpret the appearance question? *When you look at this with your eyes right now, what does it look like?* This question likely seems to the PAR subject to be a question about perceptual access. The appearance question is a question about perception for both the adult using BR and the 4-year-old using PAR. Since there is no representation mediating between the person and the world for the PAR child but only access

[52] The explanation of PAR children’s performance on A/R identity tasks from this section was inspired by conversations with Bill Fabricius.

to the world or lack of access to the world, then the PAR child must interpret this as a question about perceptual access. Now they can compute PAR:

Rule 1: Seeing → Knowing (and Not Seeing → Not Knowing)[53]

Rule 2: Knowing → Getting it Right (and Not Knowing → Getting it Wrong)

A child using PAR should reason that I'm not touching the sponge *right now* (recall that the appearance test question includes the locution 'right now'), so I don't know it's a sponge, and I'll get it wrong. "Getting it wrong" in this case means thinking that it's a rock when it's really a sponge, and so the PAR subject should respond to the appearance question with "a rock." As with the False Belief Task and the A/R Property Task, the PAR child gives the correct response but means something very different by it than the typical adult does.[54]

4. Conclusion

We have demonstrated that the major studies of children's understanding of perception are consistent with the PAR hypothesis. First, we noted that in Flavell's influential model of perception understanding (Flavell 1974), the acquisition of the concept of mental representation is what distinguishes Level 1 and Level 2. Level 1 involves understanding that the same object may be visible to one person but not to another. Hence this is a test of

[53] Remember that 'seeing' in the PAR Rules simply means *perceptual access* and may refer to any of the five sense modalities. See Hedger & Fabricius 2011 and Chapter 2 for more about the PAR rules, and especially Chapter 5 for more about interpreting what the PAR child means by 'see' and 'know' here.

[54] One way to test this hypothesis would be to give children a modified A/R identity task with three options, similar to the one used by Fabricius and Khalil (2003). In their version they showed the child that a doll (Yoshi) rode on the rock-sponge as if it were a car. In this way we could add a third option to the appearance test question: "When you look at this with your eyes right now, what does it look like? Does it look like a rock, or look like a sponge, or look like a car?" The child using PAR should divide their answers between "a rock" and "a car" but never answer with "a sponge."

perceptual access, which corresponds to the first rule of PAR.[55] Level 1 is thus required in order to use PAR, so children in this stage of theory of mind development (roughly 4- and 5-year-olds) should have no difficulty with Level 1 tasks. As we saw in Section 2 above, they do not have difficulty with these tasks.

Lacking a concept of mental representation, children in the PAR stage should fail Level 2 tasks, since Level 2 involves understanding that the same object can appear differently to different people (i.e. that two subjects can have different mental representations of the same object). Success rates on Level 2 tasks initially appear to vary widely, but we are able to dispel most of the confusion once we distinguish those tasks which putatively test Level 2 but are actually testing Level 1 understanding from those tasks truly testing Level 2. For example, a subject may use simple physical cues in order to correctly move the turntable display in Borke (1975)'s task, without considering the mind of the other person at all (e.g. Grover's car is close to the horse but far away from the house, so I'll turn the table until the horse is directly in front of me and the house is far away).

Once we are able to ignore the tasks that are actually only testing Level 1, we find that children in the PAR stage do fail Level 2 tasks for the most part, although they may be successful on some very simple Level 2 tasks, such as the Turtle task of Masangkay et al (1974). However, we have to remember that predicting the visual representation of another person is not a simple all or nothing matter, but straightforwardly admits of degree. It involves abilities

[55] More specifically, we might say that determining perceptual access is figuring out which conditional version of the first rule of PAR to compute (If S has perceptual access to X, then S knows P about X; or If S does not have perceptual access to X, then S does not know P about X).

such as simulation in addition to understanding mental representation, for instance. What's more, the transitional PAR stage presumably involves a gradual acquisition of the concept of mental representation, or at least a gradual acquisition of the knowledge and understanding required to have such a concept.[56]

The existence of a physical model, along with only two possibilities, may assist children at the beginning of this spectrum in passing the Turtle task. The child need only look at the physical picture and determine its orientation to the experimenter's line of sight. In this way, although it is most plausible to consider the Turtle task a true test of Level 2 (for reasons given in Section 2 above), the task demands are only slightly greater than Level 1 tasks such as the Picture task. They would likely fail a more sophisticated version of the Turtle task where there are four possible positions of the experimenter and subjects needed to choose between perhaps eight photographs; and indeed they do fail similar such tasks, such as the Pointing task of Fishbein et al (1972). Adults and older children who are using BR would presumably have

---

[56] Although my hypothesis is that BR is innate in the sense that it is not learned through a rational process (See Chapter 6 below; cf. Fodor 1975 on this sense of *innate*), I nonetheless believe that BR acquisition is a gradual rather than a sudden process. These two aspects are prima facie consistent with each other; this is similar e.g. to the view many cognitive scientists have about children and language (e.g. Pinker 1994).

To be clear, once a child acquires the concept of mental representation, then that child is in the BR stage. The understanding of mental representation is the dividing line between the PAR and BR stages of cognitive development. Of course, any person's understanding about mental states comes in degrees, and it's unlikely anyone has a full understanding of what it is to perceive something, for example. But someone needs to possess the concept PERCEIVE before they can think about it at all, and children in the PAR stage do not share the adult concept SEE e.g., because their idea of seeing is non-representational. They do have some concept SEE, but it is a different concept from the adult concept. See Chapters 5 and 6 below for more about PAR versus BR concepts.

One process that needs more research is exactly how the transitions between the three ToM stages work. More longitudinal studies of ToM development would be useful. To be precise, it is an open empirical question exactly how long it takes to acquire the adult BR concepts. Although I assume it is somewhat gradual, it is of course relatively quick (even if it take 2-3 years) given the breadth of knowledge required to do folk psychology. In this sense it is much like acquiring the ability to talk.

high success rates on such a task, however. For instance, Fishbein et al (1974) found that 8-year-olds chose the correct photograph (out of eight options) 74% of the time on the Pointing task. All things considered, the data on Level 1/ Level 2 tasks appears consistent with the PAR hypothesis.

Mental states such as perceptions are of course representation*s.* Furthermore, as philosophers initially suggested to psychologists (Bennett, 1978; Dennett, 1978; Harman, 1978), a good way to test whether someone understands mental representation is to test whether they are sensitive to *misrepresentation*, as is done in the false belief task. It would be implausible to say that someone can understand representation without realizing that representations can sometimes misrepresent reality. John Flavell and his colleagues were the first to investigate when children acquire the understanding that things can be different than they appear to us (Flavell et al, 1983), which is an important part of recognizing that sensory states are representational, and hence may occasionally misrepresent reality. According to the PAR hypothesis, children don't understand belief until after about six years of age, and if the lack of understanding of mental representation is consistent across multiple tasks, then the appearance/ reality distinction should likewise not be acquired until this stage of theory of mind development.

In Section 3 we demonstrated that once we realize how the test questions of appearance/ reality tasks must be interpreted by children using PAR, the findings on these tasks are consistent with the PAR hypothesis. Also, the PAR theory helps explain the findings of Flavell et al (1989), Russell and Haworth (1988), and Braine and Shanks (1965a, 1965b) which have been puzzling to researchers of cognitive development. The evidence from studies of children's understanding of perception can thus be added to the case for the PAR hypothesis, and the view that children don't acquire the concept of mental representation until fairly late in their development, i.e. typically around six years of age, contrary to every other view of cognitive development in the literature.[57]

[57] As an homage to the late, great Jerry Fodor, we could nickname this monograph *A brief refutation of developmental psychology.*

References

Aristotle, *De Anima.* In *The Complete Works of Aristotle: The Revised Oxford Translation, Volume 1*, Trans. J. Barnes, Princeton University Press, 1994.

Astington, J. W., Harris, P. L., & Olson, D. R., Eds. (1988) *Developing theories of mind.* Cambridge, England: Cambridge University Press.

Baillargeon, R., & DeVos, J. (1991). Object permanence in young infants: Further evidence. *Child Development, 62,* 1227-1246.

Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences, 1,* 557-560.

Borke, H. (1975). Piaget's mountains revisited: Changes in the egocentric landscape. *Developmental Psychology, 11,* 240-243.

Braine, M. D. S., & Shanks, B. L. (1965a). The development of conservation of size. *Journal of Verbal Learning and Verbal Behavior, 4*, 227-242.

Braine, M. D. S., & Shanks, B. L. (1965b). The conservation of a shape property and a proposal about the origin of conservations. *Canadian Journal of Psychology, 19*, 197-207.

Carey, S. (2009). *The origins of concepts.* Oxford, England: Oxford University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: The MIT Press.

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences, 1*, 568-570.

Descartes, R. *Meditations on First Philosophy.* In *The Philosophical Writings of Descartes, Volume II,* Trans. J. Cottingham, R. Stoothoff, and D. Murdoch, Cambridge University Press, 1985.

Fabricius, W. V., & Khalil, S. L. (2003). False beliefs or false positives? Limits on children's understanding of mental representation. *Journal of Cognition and Development, 4*, 239-262.

Fishbein, H. D., Lewis, S., & Keiffer, K. (1972). Children's understanding of spatial relations: Coordination of perspectives. *Developmental Psychology, 7*, 21-33.

Flavell, J. H. (1974). The development of inferences about others. In T. Mischel, Ed., *Understanding other persons.* Oxford: Blackwell, pp. 66-116.

Flavell, J. H. (1986). The development of children's knowledge about the appearance-reality distinction. *American Psychologist, 41:4,* 418-425.

Flavell, J. H., Botkin, P. T., Fry, C. L., Wright, J. W., & Jarvis, P. E. (1968). *The development of role-taking and communication skills in children.* New York: Wiley.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1—Level 2 distinction. *Developmental Psychology, 17,* 99-103.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance-reality distinction. *Cognitive Psychology, 15*, 95-120.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1989). A transitional period in the development of the appearance-reality distinction. *International Journal of Behavioral Development, 12,* 509-526.

Flavell, J. H., Shipstaed, S. G., & Croft, K. (1978). Young children's knowledge about visual perception: Hiding objects from others. *Child Development, 49,* 1208-1211.

Fodor, J. A. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

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

Gopnik, A., & Wellman, H. M. (1994). The "theory theory." In L. Hirschfield and S. Gelman, Eds., *Domain specificity in culture and cognition*. New York: Cambridge University Press, 257-293.

Harman, G. (1978) Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences*

1, 576–77.

Hedger, J. A., & Fabricius, W. V. (2011). True belief belies false belief: Recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development." *The Review of Philosophy and Psychology, 2*, 429-447.

Kretch, K. S., & Adolph, K. E. (2013). Cliff or step? Posture specific learning at the edge of a drop-off. *Child Development, 84,* 226-240.

Laurendeau, M., & Pinard, A. (1970). *Development of the concept of space in the child.* New York: International Universities Press.

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Development, 45,* 349-372.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: The MIT Press.

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space.* London: Routledge and Kegan Paul.

Piaget, J., & Inhelder, B. (1974). *The child's construction of quantities.* Trans. A. J. Pomerans. London, England: Routledge and Kegan Paul.

Pinker, S. (1994). *The language instinct: How the mind creates language.* New York: William Morrow and Company.

Russell, J., & Haworth, H. M. (1988). Appearance versus reality in dyadic interaction: Evidence for a lingering phenomenism. *International Journal of Behavioral Development, 11*, 155-170.

Yaniv, I., & Shatz, M. (1988). Children's understanding of perceptibility. In Astington et al (1988), pp. 93-108.

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition, 35,* 41-68.

CHAPTER FOUR

Perceptual access reasoning:

Developmental stage or System 1 heuristic?

1. Introduction

In her paper "Mindreading with ease? Fluency and belief reasoning in 4- to 5-year-olds," Anika Fiebich (2014) discusses the Perceptual Access Reasoning (PAR) hypothesis (Chapter 2, Fabricius et al 2010, Fabricius and Imbens-Bailey 2000, Fabricius and Khalil 2003, Hedger and Fabricius 2011, Hedger 2016), and proposes an alternative hypothesis. The PAR hypothesis and some of the evidence in support of it was presented in Chapter 2. In Section 2 I briefly review the Fluency Hypothesis, and then point out a number of problems with Fiebich's proposal and argue for the superiority of the PAR hypothesis in Section 3. In the conclusion I explain that the PAR hypothesis does leave room for a dual-systems theory of sorts, albeit one which is importantly different from Fiebich's proposal. I close by hinting at a broader implication for theorizing about child psychology which falls out of the response to Fiebich.

2. The fluency hypothesis

Fiebich (2014) agrees that PAR explains the failure of 4- and 5-year-olds on the true belief task, but disagrees that children at this stage are also using PAR to pass the false belief task.[58] She argues that children use Belief Reasoning (BR) to pass the false belief task. Her hypothesis is that children switch from BR on the false belief task, to PAR on the true belief task

[58] For details about these tasks, see Chapter 2.

because the latter is a cognitively cheaper heuristic. Adopting Kahneman's well-known dual systems theory (Kahneman 2011), she argues that fluency explains when children ages 4 and 5 use BR or PAR. When the cognitive task feels easy (e.g. the true belief task), they default to the fast and frugal PAR heuristic, and when the task feels more difficult (false belief task), they use the more effortful and deliberative BR.

In the third section of her article, Fiebich (2014) discusses at length the cognitive demands involved in each of the three strategies which, according to the PAR hypothesis, are used during the three distinct stages of ToM development (Chapter 2, Hedger and Fabricus 2011). Children under the age of 4 tend to fail verbal false belief tasks. In the false location task, they tend to choose the green cupboard, because that is where the chocolate is in reality (disregarding Maxi's mental representations), and so Hedger & Fabricus (2011) refer to this as *Reality Reasoning (RR). Perceptual Access Reasoning (PAR)* is the use of the two PAR rules to interpret and predict behavior, and they dub a full-fledged theory of mind including an understanding of belief *Belief Reasoning (BR).*

Fiebich (2014) argues convincingly that "Perceptual Access Reasoning is cognitively more demanding than Reality Reasoning but less demanding than Belief Reasoning" (p. 935). Kahneman (2011)'s view is that people tend to use cognitively cheaper System 1 reasoning processes, unless some cognitive strain (i.e. feeling of difficulty) induces them to use the more effortful but also more reliable System 2 reasoning processes. Adopting this theory, Fiebich proposes that children ages 4 and 5 use the System 1 heuristic PAR on the true belief task, but that the cognitive strain involved in the false belief task causes them to use System 2 BR.

The explanation of this cognitive strain relies on another theory from psychology, viz. cognitive dissonance (Festinger 1957). Normally, cognitive dissonance in social psychology refers to the discomfort felt by a person who holds two or more conflicting beliefs (or values or emotional responses). That is, cognitive dissonance is normally experienced when one realizes that two of her own beliefs logically conflict with each other. In her paper, Fiebich proposes that the false belief task feels more difficult because of cognitive dissonance resulting from the child's own true belief conflicting with the false belief *attributed to Maxi* (the protagonist in the belief task story). In her words, "4- to 5-year-old children experience cognitive dissonance in reasoning processes in false belief tasks in which their own belief… differs from that of the agent but not in true belief tasks in which there is no such difference" (p. 941). After a couple of years, the "repeated experience and learning" of using BR under the cognitive strain of false belief situations allows children to "make use of BR with ease" (p. 941), and use it in true belief situations.

3. Problems for the fluency hypothesis

The empirical support Fiebich relies on is Perner and Horn (2003)'s apparent failure to replicate the findings of Fabricius and Khalil (2003) on 3-location false belief tasks.[59] In the next subsection I present five reasons to question the findings of Perner and Horn (2003), before directly addressing the fluency hypothesis in the two subsections which follow. The main

---

[59] Of course, according to Perner and Horn (2003) children use BR consistently at this stage, and according to Fabricius and colleagues children use PAR consistently at this stage, while Fiebich's idea contains a switching back and forth between the two reasoning strategies, depending upon the task. I am tempted to appeal to parsimony here, but I'm also aware that these sorts of debates all too often become an irreconcilable conflict of intuitions, and so I set that worry aside for the remainder of the chapter.

problem with the fluency hypothesis is that it plucks PAR out of the home background theory in which it makes sense when considering all of the data, and thrusts it into various background theories taken from adult social psychology. Not only does PAR not fit at all neatly into the theories of Kahneman (2011) and Festinger (1957), as demonstrated in section 4.2, but changing PAR from a developmental stage into a temporary heuristic causes some logical troubles for the fluency hypothesis, which are explained in Section 4.3.

3.1 A closer look at the findings of Perner and Horn

First, the Perner and Horn (2003) explanation of the Fabricius and Khalil (2003) data, which Fiebich (2014) agrees with (p. 933), is that children in the Fabricius and Khalil study were confused by a series of three yes/no questions about the agent’s (and their own) beliefs. This confusion, rather than using PAR, is supposed to account for the subjects’ poor performance on the 3-location false belief task. However, if children were genuinely confused about the questions, then one would expect them to switch their answers to the series of questions across tasks. Fabricius and Khalil (2003) reported that the reality responses were consistent across tasks, which casts serious doubt on the confusion explanation.

Second, the percentages of children passing the control questions in Perner and Horn (2003) are alarmingly low—less than 75% in each of the two studies. In Wellman et al (2001)’s meta-analysis of false belief studies, the findings of Perner and Horn wouldn’t have been included in their “primary conditions,” because they required that more than 80% of children pass the control questions. In contrast, Fabricius and Khalil (2003) reported that in 3-option tasks 5- and 6-year-olds were almost always well in excess of 90% correct on control questions.

Third, although Perner and Horn (2003)'s findings on the 3-option *location* and *neutral box* false belief tasks were contrary to the PAR hypothesis, their findings on the 3-option *contents* false belief task (which they call the "typical box" task) actually confirm the PAR hypothesis, which predicts that 4-year-olds should choose randomly between the false belief and irrelevant options. In Study 2, six subjects chose the false belief option, but three chose the irrelevant option, with distribution not significantly different than chance.[60]

Fourth, a meta-analysis (Wellman et al 2001) found that across 178 false belief studies, rates for passing the two option location task are highly consistent with those for passing the typical contents task. Perner and Horn (2003, Study 2) failed to replicate this most basic finding in the false belief literature, and found instead that their location tasks were much easier; children passed them at a rate of 75%, as contrasted with only 38% on the contents tasks.[61] These anomalous findings suggest that there is a problem with the procedures used by Perner and Horn (2003)[62].

Fifth, Fabricius and Khalil (2003) argue that differences across tasks in the relative salience of the three options may account for not only the PAR-contrary findings but also the

---

[60] The *"typical box" task* of Perner and Horn (2003) followed the Unexpected Contents or "Smarties" false belief task described in Chapter 2, except that in the 3-option version the protagonist removes the pencil and exchanges it with a pebble before asking the test question. This was similar to the 3-option contents task used by Fabricius and Khalil (2003), and in both studies some subjects chose the irrelevant option (a pencil in this particular case), which is consistent with what the PAR hypothesis would predict. (You will recall that a child using PAR should reason that Elmo doesn't see what's inside the container, and therefore won't know what's in the container and will "get it wrong." Thus, given a forced choice she should choose randomly between the false belief contents and the irrelevant contents, since both options are incorrect.) In the *neutral box* task of Perner and Horn (2003), a plain box with no markings or color was used in place of the familiar candy container.

[61] Out of 21 subjects, 14 passed the two option location task and 8 passed each of the two versions of the two option contents tasks (for a total of 16 out of 42). Subjects also found the three option location task to be much easier than the three option contents tasks.

[62] Perner and Horn (2003) admit that they don't have an explanation for this anomaly (p. 269).

anomalous finding noted above. In order to focus the child on each option in turn and thus minimize the effects of salience differences, Fabricius and Khalil (2003) asked a forced choice test questions for each alternative (e.g., “Will Maxi think it is in the [reality location]?” “Will he think it is in the [irrelevant location}?” “Will he think it is in the [false belief location]?”). Perner and Horn (2003) asked one open-ended test question (e.g., “Where will he think it is?”), requiring the child to consider all the alternatives at once and thus exposing a child using PAR to the influence of salience differences. In location tasks all the options are present, but in both contents tasks only the one box is present and it has the reality object inside, so without the questions reminding the child of the other options they could easily forget them, which would explain the anomalous finding that there were more choices of the reality option in the contents tasks. In addition, in the location and neutral box tasks, but not in the typical box task, the fact that the protagonist chose and thereby expressed a preference for the false belief option could have given that option extra salience, which would explain why Perner and Horn found significantly fewer choices of the irrelevant options in only those tasks.

In sum, the putative explanation that confusion on the part of the subjects is the reason for their poor performance on 3-location tasks in Fabricius & Khalil (2003) looks to be implausible since the children gave consistent answers to all of the questions across tasks, suggesting that they were not confused. Furthermore, the findings of Perner & Horn (2003) ought to be met with skepticism on independent grounds—viz. subjects’ poor performance on the control questions, and an anomalous discrepancy between their performance on location and contents tasks—which suggest that there may be problems with the tests they used. Also, Perner and Horn’s (2003) results on the 3-option contents false belief task are consistent with the predictions of the PAR hypothesis. Finally, Fabricius & Khalil's (2003) hypothesis that the one open-ended question in Perner and Horn's (2003) procedures (which fails to remind subjects of all three locations or objects) resulted in salience differences across tasks, is able to explain the results which appear to contradict the PAR hypothesis as well as the anomalous inconsistency between location and contents task results.

Thus there is one criticism of the Fabricus and Khalil (2003) procedure (viz. that three yes/ no questions may have been confusing to subjects), which is endorsed by both Perner and Horn (2003) and Fiebich (2014). Here I have presented evidence against that criticism, as well as four other inter-related reasons for doubting the Perner and Horn (2003) findings.

3.2 PAR doesn't fit neatly into Kahneman's background theory

Even if we disregard the issues with Perner and Horn (2003), however, there are serious concerns for Fiebich's proposed explanation of ToM findings. First, it is not at all obvious that the false belief task is more difficult or causes more cognitive strain than the true belief task, in the relevant sense of placing further task demands on the subject. Fabricius et al (2010) argue that the task demands of the true belief tasks are as similar as possible to the original false belief tasks on which they are based, and Fiebich provides no reason to think otherwise. The cognitive dissonance explanation is a non-starter. There is no evidence that holding a belief which conflicts with a belief *attributed to someone else* causes cognitive dissonance in adults, despite the vast literature on this subject. For one thing, this seems to be an all-too-common occurrence.[63] Furthermore, there is no evidence to suggest that cognitive dissonance plausibly creates the kind of cognitive strain which typically cause subjects to use System 2 reasoning. In addition to these difficulties, we would also need some evidence that cognitive dissonance of any kind exists in children, and that it is similar to that of adults. We know very little about how these sorts of mechanisms develop.

In general, characterizing PAR as part of Kahneman (2011)'s System 1 seems to be a poor fit. Kahneman (2011) says that System 1 processes "operate automatically and quickly, with little or no effort and no sense of voluntary control" (p. 20). For Kahneman, System 1 processes are automatic and underneath the level of conscious control. He offers several examples, including recognizing anger in a facial expression (19), orienting to sudden loud sounds (22), and seeing one line as longer than the other in the Müller-Lyer illusion (26-27). However, PAR doesn't seem to be like these sorts of processes. For instance, it appears to be under conscious control since children using the PAR rules mention them in their verbal justifications for choices made on false and true belief tasks (Fabricius et al 2010), and no belief tasks (Chen et al 2015). It also seems to be an effortful process, since it takes some time to develop (Chapter 2, Fabricius et al 2010, Hedger & Fabricius 2011), and the verbal reports of subjects suggest that they undergo a kind of step-by-step deliberate reasoning process. This is in contrast to quick and automatic System 1 processes such as recognizing facial expressions or

[63] It has long been pointed out by anti-realists that we find widespread disagreement amongst professional philosophers, and yet we seem to find no discomfort or attempt to reconcile beliefs in Western philosophy's 2700 year history. The same is true in the political sphere, or at any time when people are aware of disagreeing with someone else.

perceiving lines as of different lengths, where subjects are unable to report about how they do those things or what process they used.[64]

Kahneman (2011) does say that "other mental activities become fast and automatic through prolonged practice… [for example] learned skills such as reading and understanding nuances of social situations" (22).[65] However, as Fiebich agrees, 3-year-olds typically use RR, and it's not until ages 4 and 5 that children typically use PAR (Fabricus et al 2010, Fabricus & Khalil 2003). Thus, they presumably haven't used it long enough for it to become fast and automatic.

3.3 Logical and developmental inconsistencies in the fluency hypothesis

Even more troubling, however, is that the cognitive dissonance story reverses the order of explanation. According to Fiebich, the cognitive dissonance caused by holding a belief which is inconsistent with that attributed to the protagonist in the false belief task causes 4- and 5-year-olds to use BR (p. 13).[66] However, the subject must use BR to attribute the dissonant belief to the protagonist in the first place. When we think this through developmentally we see

---

[64] An anonymous reviewer of Hedger (2016) points out this evidence is only suggestive and not decisive. However, I do not find the reviewer's examples convincing. For instance, it is suggested that agents can be conscious of a process which is not under conscious control, such as a knee-jerk reflex. The issue though is that although subjects can be aware of the *outputs* of these processes, they are not aware of the processing itself. For example, people are aware that they recognize faces and judge language strings as ungrammatical. However, subjects are unable to report about *how* they do it. In contrast, subjects in Fabricius et al (2010) are not only aware of the predictions about where Maxi will look (and the judgments about what Maxi knows)—i.e. the outputs of PAR reasoning—they are also aware of the steps of the reasoning itself, and the process by which they arrive at those predictions and judgments.

[65] He also says in Chapter 5 that repeated experience is a cause of cognitive ease.

[66] I strongly disagree with this manner of speaking, because I don't think a child using RR or PAR thinks or attributes anything about *beliefs*; but I'll reluctantly adopt it throughout the rest of this paragraph for ease of explanation. However, if the PAR hypothesis is correct (and children at this stage aren't reasoning about beliefs), then this provides another reason to reject the dissonance story.

additional problems. For instance, when a 3 –year-old uses RR on the false belief task, she attributes to the protagonist a belief about where the chocolate is *which is exactly the same as her own*, since it's just the reality location; hence, no cognitive strain. Therefore Fiebich needs some explanation of how cognitive strain ever arises on the false belief task. Notice further that the 4- or 5-year-old who uses PAR on the true belief task ends up attributing a belief which is *not* consistent with her own, since she predicts that the protagonist will "get it wrong" and not go to the reality location where the chocolate is currently located. Thus, there is cognitive strain on the true belief task, but according to Fiebich this task isn't supposed to cause cognitive strain. That's supposed to explain why subjects use PAR in the first place.

All of this just leads to deeper troubles for the fluency hypothesis. Recall that although PAR is less effortful than BR, it is also *more* effortful than RR. So now Fiebich requires an explanation of why children should use the PAR heuristic on the true belief task when they already have a cognitively cheaper heuristic, viz. RR. What's more, *RR gives the correct result on the true belief task*, contrary to PAR. If fluency were the sole reason children ages 4 and 5 revert to PAR on the true belief task, then why wouldn't they instead revert to RR? It's both cheaper and better than PAR on the true belief task.

In fact, if PAR is only ever used in true belief situations, where it gives the wrong result, how and why would it ever be used in the first place? If 3-year-olds already use a successful heuristic, then why develop PAR which is both more cognitively demanding and is also unsuccessful when used for the sole purpose it is designed for, according to Fiebich's hypothesis? She says, "Note that I have not made any claims about how RR is replaced by PAR in younger children" (p. 14). However, this is needed on her story. For one thing, an explanation would seem to be required for how such a mechanism as Fiebich's PAR should ever persist through the natural selection process, given that it always gives the wrong result. Note that on the PAR hypothesis, this feature is merely a spandrel piggybacking on the fact that PAR gives the correct result in false belief situations (Chapter 2, Hedger & Fabricius 2011). According to Fiebich, however, PAR isn't used on false belief tasks. In fact, according to her story, children develop PAR and BR at the same time. Hence we lack an explanation of why PAR ever develops in the first place, given (a) that a cognitively cheaper and more successful strategy is already in place for true belief situations (namely RR), (b) according to the fluency hypothesis a new strategy (namely BR) is simultaneously available which is a successful predictor of true or false *beliefs*, and (c) PAR is *never* successful according to Fiebich's account.

4. Conclusion: Reflections and a more plausible dual systems theory

Although I believe that Fiebich's particular proposal has been shown to be implausible, the PAR hypothesis does make room for a dual systems analysis. Recall from Chapter 2 that Hedger and Fabricius (2011) conjecture that a developmental precursor of PAR, which they dub *Rule A*, may be at work as an implicit theory of mind mechanism. The two PAR rules (**see/not see → know/not know** and **know/not know → get it right/get it wrong**) are bridged by the concept KNOW, which children begin to acquire and link with perceptual access by about 3 ½ years of age (e.g., Pillow, 1989; Pratt & Bryant, 1990). Rule A is the condensed rule **see/not see → get it right/get it wrong**, which Hedger & Fabricius (2011) hypothesize children implicitly use before they acquire the concept KNOW.[67] Rule A (like PAR) should cause subjects to pass the false belief task but fail the true belief task. Although they operate similarly (e.g. leading to the same prediction in a false belief task), PAR and Rule A are nonetheless distinct cognitive mechanisms. Rule A is implicit, and modular to some important degree. Thus for instance the output of Rule A is not available to explicit reasoning for 3-year-olds (Clements & Perner, 1994). PAR is an explicit, conscious reasoning process, during which children use the word "know" in their verbal explanations (Fabricius et al, 2010; Chen et al, 2015).

The Rule A hypothesis could explain the data that 3-year-olds give the incorrect verbal response during the two-option false belief task but also show anticipatory looking to the correct location (Clements & Perner, 1994; Garnham & Perner, 2001; Garnham & Ruffman, 2001; Ruffman, Garnham, Import, & Connelly, 2001). It is also consistent with the studies of infant and chimpanzee theory of mind that they are using Rule A (see Hedger & Fabricius 2011 for more details). Hedger and Fabricius (2011) conjecture that Rule A may also persist into adulthood. In other words, Rule A may exist simultaneously in adult cognition along with BR, along the lines of a Dual Systems model such as Kahneman (2011). This could be tested by using the "eye gaze" methodology developed for children under 4 years of age (Clements & Perner, 1994).[68] Hence the hypothesis is that Rule A is a non-verbal cognitive system used throughout

---

[67] Of course, the concept used by children in the PAR stage of development is not the same concept of knowledge used by adults, since the former is non-representational. More on this topic in the next two chapters.

[68] However, in order to know whether correct anticipatory looking in the false belief task indicates attribution of false beliefs or use of Rule A, the methodology needs to include a true belief task in which there is some interruption in the agent's connection to the situation that is comparable to what occurs in the false belief task (i.e. an appropriate cue for situation change). The previous eye gaze studies have not included such true belief tasks (Hedger & Fabricius, 2011). See also Chapter 7, Sections 2.2 and 2.3.

human development and present in chimpanzees. More research into the anticipatory looking of children and adults would help to disconfirm or point in favor of the Rule A conjecture.

If adults were found to continue to use Rule A implicitly it could explain a puzzling feature of theory of mind—that at times we appear able to make judgments about the mental states of others quickly, automatically, and effortlessly, while at other times the process is difficult and deliberative. There is evidence that BR is effortful and difficult for adults (Apperly et al, 2006; Keysar et al, 2003). Lin, Keysar and Epley (2010) found that higher working memory capacity can have a positive impact on adult performance in theory of mind tasks, while cognitive load impairs this ability. Perhaps this can be explained by BR and Rule A being different psychological mechanisms, the former explicit and effortful and the latter implicit, automatic and modular[69]; System 2 and System 1 in Kahneman (2011)'s terminology. Thus, even though I argue against Fiebich's specific account of a dual-systems theory (2014), the PAR hypothesis does make room for a dual-systems account in which an efficient but limited and rigid cognitive ToM mechanism (Rule A) exists alongside a more flexible but cognitively demanding cognitive ToM mechanism (RR, PAR or BR, depending upon the developmental stage of the subject).[70]

[69] On modularity and psychology, see Fodor (1983) and Barrett & Kurzban (2006).
[70] Although the sketch of a dual-systems account presented here is superficially similar to the one proposed by Fiebich (2014), they are importantly different for at least four reasons: According to my proposal (but not Fiebich's) (1) Rule A and PAR are distinct psychological mechanisms, (2) when using a conscious process in a psychological test such as the verbal false belief task, subjects will use the effortful, deliberative reasoning process, (3) this latter process is different depending upon which stage of theory of mind development a subject is in—RR, PAR, or BR, and (4) the features that might make a BR subject revert to Rule A would be perhaps time constraints or anticipatory looking procedures, but not features of a particular task. Thus, 3-year-olds fail the 2-location false belief task when tested using a verbal report method (by using RR), but pass using an AL measure (by using Rule A).

In fact, the Rule A hypothesis is compatible with other arguments for dual-systems accounts of belief reasoning, such as that presented by Apperly and Butterfill (2009). In this excellent paper, Apperly and Butterfill (2009) point out that adults face competing demands upon their ToM abilities—they must at some times be fast and efficient, but at other times be flexible and reliable (pp. 953, 956-957). Both demands could be satisfied by utilizing different cognitive mechanisms for different purposes. They also argue that a dual-systems account could explain the prima facie paradoxical findings that infants pass some false belief tasks while older children fail others (pp. 957-958; see also De Bruin & Newen 2014). This is explained by the fact that infants and toddlers are using distinct cognitive mechanisms (for more on how the PAR theory can explain this paradox, see Chapter 2 and Hedger & Fabricius 2011). They also point out that the ToM abilities of infants are quite commensurable with those of non-human animals such as chimpanzees, suggesting that they may be using the same lower-level mechanism (p. 958). As we saw in Chapter 2, Rule A may be the mechanism which both chimps and infants are using (see Table 1). Hence the PAR/ Rule A hypothesis may be seen as simply specifying the more general dual-systems account of Apperly & Butterfill (2009), and arguments which they present also lend support to the present hypothesis.[71]

In sum, according to the PAR hypothesis, infants pass VOE versions of the 2-option false belief task by using Rule A, which is presumably a modular or System 1 cognitive process. 3-year-olds fail the verbal 2-option false belief task, because they are using the conscious, System 2 reasoning process known as Reality Reasoning. When children enter the next developmental

[71] The PAR/ Rule A hypothesis also seems to be consistent with the general considerations raised by De Bruin and Newen (2014), although I am unsure about the specifics of their Association Module/ Operating System account.

stage around the age of four, and begin to use Perceptual Access Reasoning consciously and deliberately, they again pass the (verbal) 2-option false belief task. This same pattern of development which is part of the PAR hypothesis is also seen in studies of object solidity. 4-month-olds pass VOE measures of object solidity, which Carey (2009) hypothesizes is accomplished by using a modular cognitive-perceptual process. 2- and 3-year-olds fail searching versions of the same task, because they are using a different effortful and conscious process. By the age of four they pass the searching measures of object solidity, which Carey (2009) hypothesizes is because the new deliberate reasoning process now incorporates the same mental representations which were already being used inside the modular perception process (see Chapter 2 for more details).

Fiebich's proposal has been shown to have numerous difficulties which make the fluency hypothesis problematic. Note that Fiebich eventually gets backed into this corner by accepting the findings of Perner and Horn (2003) and denying the findings of Fabricius & Khalil (2003). The acceptance that children use BR on false belief tasks must be reconciled with Fabricius et al.'s (2010) findings that children at this stage fail true belief tasks, and Fiebich attempts to do so by taking the position that children switch back and forth between two different reasoning strategies. Note that others (Friedman et al 2003, Lagattuta et al 2010, Ruffman 1996) have claimed that when children fail true belief tasks, and fail false belief tasks by choosing the irrelevant option, they are nevertheless attributing false beliefs to those protagonists. Fabricius et al (2010) pointed out that such a position blurs the "ordinary concept of belief [which] requires that it is caused by some information that justifies the believer in taking it as true" (p. 1407). Fiebich takes a more coherent position, but the fluency hypothesis eventually comes up short because it is internally inconsistent as well as a poor fit with the theories of cognitive dissonance and dual systems cognition.[72]

A crucial aspect of the sort of inference to the best explanation done by scientists (including psychologists) is a background theory which is not only consistent with, but also helps to explain, the data. The PAR hypothesis is grounded in a developmental theory. The theory is consistent with the set of general beliefs accepted in childhood developmental psychology. Hedger and Fabricius (2011) further demonstrate the compatibility of the PAR hypothesis with recent ToM findings in chimpanzees and human infants (see also Chapter 2). Fiebich's position is logically possible; children could switch strategies. However, when we look closer we find numerous difficulties. Researchers need to take note that incorporating processes at work in adults, without considering how they develop or fit into background theories, is likely to cause more problems than it solves for explanations of childhood cognition.

---

[72] As an anonymous reviewer points out, Fiebich (2014) proposes an empirical test of her Fluency hypothesis (pp. 941-942). I believe that the reasoning in Section 4 of this paper is sufficient to show that the hypothesis is implausible, but even if one disagrees with that, there are other problems with the first of Fiebich's experimental paradigms. First she suggests testing whether 4- and 5-year-olds experience more cognitive strain in the 2-location false belief task than in the true belief task of Fabricius et al (2010). The difficulty here is that, first, I'm not sure that there are accepted objective signals of cognitive strain in general, much less in children. Second, even if we did find more cognitive strain in the false belief task, that would not in any way count as evidence against the PAR hypothesis.

Fiebich also suggests adding variables that induce cognitive strain to the true belief task in order to test whether that allows 4- and 5-year-olds to pass. Again, even if this test confirmed her hypothesis I'm not sure that would count as evidence against PAR, but I do accept that it would perhaps lend support to the Fluency Hypothesis, *if* the internal inconsistencies of the theory could somehow be resolved. If we indeed found that making a task more difficult improved the performance of 4- and 5-year-olds, then that would seem to at least be evidence for a 2-systems account of some kind or other.

[This chapter is a revision of material which originally appeared in Hedger (2016). Thanks to Bill Fabricius, Bob Van Gulick and Kevan Edwards for comments on earlier drafts of this chapter. Thanks also to Jesse Prinz for a helpful conversation about this chapter. Portions of this chapter were also presented at the Social Minds conference at the University of Maryland, College Park (April 2015). Thanks to participants who gave comments and feedback, in particular Tad Zawidzki.]

References

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

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17,* 841-844.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review, 113,* 628-647.

Carey, S. (2009). The origin of concepts. Oxford, England: Oxford University Press.

Chen, Y., Su, Y., & Wang, Y. (2015). Young children use the "ignorance = getting it wrong" rule when predicting behavior. *Cognitive Development, 35,* 79-91.

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

De Bruin, L. C., & Newen, A. (2014). The developmental paradox of false belief understanding: A dual-system solution. *Synthese, 191,* 297-320.

Fabricius, W. V., Boyer, T., Weimer, A.A., & Carroll, K. (2010). True or False: Do five-year-olds understand belief? *Developmental Psychology*, *46*, 1402-1416.

Fabricius, W.V., & Imbens-Bailey, A. L. (2000). False beliefs about false beliefs. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning about the mind* (pp. 267-280). Hove, England: Psychology Press.

Fabricius, W. V., & Khalil, S. L. (2003). False beliefs or false positives? Limits on children's understanding of mental representation. *Journal of Cognition and Development, 4*, 239-262.

Festinger, L. (1957). A theory of cognitive dissonance. Stanford: University Press.

Fiebich, A. (2014). Mindreading with ease? Fluency and belief reasoning in 4- to 5-year-olds. *Synthese, 191,* 929-944.

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

Friedman, O., Griffen, R., Brownell, H., & Winner, E. (2003). Problems with the seeing = knowing rule. *Developmental Science, 6*, 505–513.

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

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.

Hedger, J. A. (2016). Perceptual access reasoning: Developmental stage or System 1 heuristic? *Phenomenology and the Cognitive Sciences, 15,* 207-226.

Hedger, J. A., & Fabricius, W.V. (2011). True belief belies false belief: Recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development. *Review of Philosophy and Psychology, 2,* 429-447.

Kahneman, D. (2011). Thinking, Fast and Slow. London: Penguin Books.

Keysar, B., Lin, S. & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25-41.

Lagattuta, K. H., Sayfan, L., & Blattman, A. J. (2010). Forgetting common ground: Six- to seven-year-olds have an overinterpretive theory of mind. *Developmental Psychology, 46,* 1417-1432.

Lin, S., Keysar, B., & Epley, N. (2010). Refelexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology, 46,* 551-556.

Perner, J., & Horn, R. (2003). Knowledge or false negatives: Do children of 4 to 5 years simulate belief with "not knowing = getting it wrong?" *Journal of Cognition and Development, 4*, 263-273.

Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology, 47,* 116–129

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development, 61,* 973–982.

Ruffman, T. (1996). Do children understand the mind by means of simulation or a theory?

Evidence from their understanding of inference. *Mind and Language, 11,* 388-414.

Ruffman, T., Garnham, W., Import, A., & Connelly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology, 80*, 201-224.

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.

CHAPTER FIVE

Answering the Perner-Povinelli Challenge:

Distinguishing Mentalist from Behavioral Rules in Children's Theory of Mind

1. Introduction

After three decades of research, the false belief task, pioneered by Wimmer and Perner (1983), remains the most popular test of theory of mind (ToM). At this mature stage of research on ToM, many psychologists feel justified in taking a sanguine view of putatively well-established facts based on putatively well-established methods. Henry Wellman, one of the pioneers of the ToM field, and his colleagues recently write (Wellman et al 2011)…

> The demonstration of robust, early-developing theory-of-mind achievements in children worldwide – in traditional and nontraditional, western and nonwestern societies (e.g. Wellman et al 2001) – has led to tremendous interest in the ontogenetic and evolutionary origins of theory of mind. …
>
> Our measure of theory of mind is a battery of false belief tasks. *Explicit* false-belief understanding [emphasis added] is a milestone, universal theory-of-mind achievement of the preschool years (Wellman et al 2001), and is the most commonly used measure in research examining individual differences in theory of mind during the preschool years...

In Chapter Two I dubbed this the Traditional View of ToM development. It was the accepted view for the first two decades of ToM research on human children.

However, there has been recent evidence that infants and chimpanzees might pass false belief tasks (see Chapter 2). This poses a serious challenge to the Traditional View. Perhaps the view could be saved by the hypothesis that chimps and infants are passing false belief tasks by using behavioral rules instead of a concept of mind. However, this way out of the puzzle poses a new challenge for the Traditional View: If chimps and infants are passing false belief tasks by using behavioral rules, then so might preschoolers. In some cases, the sanguine view seems to be preserved by a creeping nativism. Thus Josef Perner, another ToM pioneer, recently writes (Perner 2010):

> Behavior rules capture the causal dependence of [actions] on [situations] without saying anything about the mind that creates those dependencies. But – and this is important – what they capture reflects the working of the mind. Behavior rules provide an *implicit* theory of the mind– that is, they capture the workings of the mind without mentioning the mind…
>
> Mentalism tries to capture the same S[ituation]-A[ction] dependencies but by being *explicit* about the intervening mind.

This brings me to the current status of ToM research. Most of the theory of mind establishment continue to argue that 4 ½ is the crucial age for understanding representational mental states such as belief, and most of the argument is against a resurgence of the Nativist View which has been prompted by recent infant studies in the wake of Onishi and Baillargeon (2005). In Chapter Two I demonstrated that the Perceptual Access Reasoning (PAR) hypothesis is able to reconcile all of the existent data under one comprehensive theory of ToM development (see also Hedger 2016, Hedger & Fabricius 2011). What's more, the PAR theory provides an explanation of recent findings which present troubles for both of the rival views (Traditional and Nativism), namely the data that 4 and 5-year-olds fail a true belief task (Fabricius et al 2010; see Chapter 2).

A further difficulty arises when attempting to determine which hypothesis about infant ToM is true. Cognitive scientists have recently been concerned with the aforementioned findings that chimps and infants pass versions of the false task. Psychologists worry about finding an empirical method for determining whether chimps and infants are passing ToM tasks by using mentalist rules, which include a conception of mental representation, or whether they are passing them merely by using behavioral rules which require no understanding of the mind as such. Daniel Povinelli (2001; Povinelli & Vonk, 2003, 2004; Penn & Povinelli, 2007) has made the argument that there is in fact no empirical method for deciding which of these two explanations is true, and Josef Perner (2010, forthcoming) has recently applied the argument to the case of human subjects.

In this chapter, I demonstrate how the PAR theory can provide an answer to this challenge. In fact, we already have evidence that many 4- and 5-year-olds have been using rules which make no reference to beliefs in order to pass false belief tasks, and infants and chimps may be using these same behavioral rules (or some slight variation thereof). I provide an empirical test for determining whether or not this is indeed the case, thus solving the problem. Furthermore, exploring this issue demonstrates that children under the age of six lack a conception of mental representation, and therefore lack the adult concepts of perception and belief.

2. The Perner-Povinelli Challenge

Perner (2010, forthcoming) and Povinelli (2001; Povinelli & Vonk, 2003, 2004; Penn & Povinelli, 2007) worry that in principle no empirical test of theory of mind can determine whether a given organism is using mentalist rules which refer to the agent's representational mental states, or simply using behavioral rules which merely take into account the agent's behavior in the immediate situation. Given any mental rule which makes reference to mental states, Perner (forthcoming) argues that the rule can be replaced by a "computationally equivalent" behavioral rule "by simply deleting the reference to the mental states and adjusting the rule" (p. 6).[73] The resulting behavioral rule will provide just as valid an explanation of the organism's behavior as the mental rule. The worry is that no empirical test could in principle determine whether the organism is using the mental rule or merely the behavioral version. Hence, there is no way to determine whether e.g. a child is using an understanding of the representational mind in order to pass the false belief task, or merely using (either explicitly or implicitly) a behavioral rule which produces the same behavioral output as a mental rule would. Hence there may be no empirical method for attributing a theory of mind.

Perner (forthcoming) expresses the concern this way:

> …[O]ne cannot arrive at a decisive test by demonstrating more and more appropriate responses in situations where we think we apply a theory of mind. This strategy will not help because the demonstrated response can, in principle, be based on a behavioral rule. The other issue concerns the question of how we can demonstrate the use of a theory of mind at all. In Fact: How do we know that we are using one? The existing literature has focused on plausibility arguments, in particular about parsimony. … Unfortunately, depending on how one counts assumptions both camps claimed to have parsimony on their side. (p. 6)

---

[73] A note on terminology: The behavioral rules discussed in this chapter (and by Perner and others) should not be confused with *behaviorism.* No one is claiming that the behavioral rules need to be completely devoid of intentionality. Rather, the key difference is that behavioral rules lack the conception of *mental representation*.

Perner (forthcoming) gives an example of a mental rule which would allow a person to pass the false belief task. It is composed of a *belief inference* and a *desire inference* as two input rules, and then an *action inference* which predicts the agent's behavior as a function of the mental states computed by the two input rules. The overall mentalist rule procedure for passing the false belief task then looks like the following according to Perner (forthcoming):[74]

(Belief inference): IF a person sees an object being put inside location A and then doesn't see it moved unexpectedly to location B THEN the person *thinks* the object is still in A.

(Desire inference): IF a person shows liking for an object and is heading vaguely in the direction of the object THEN he *wants* to get the object.

(Action inference): IF a person *wants* to get an object THEN the person *will go* to where he *thinks* the object is. (pp. 4-5)

By removing all of the references to mental states which appear in italics, Perner is able to come up with a behavioral rule which he claims would have the same empirical results:

(Behavior rule): IF a person sees an object being put inside location A and then doesn't see it unexpectedly moved to location B

---

[74] I would like to note the initial implausibility of the claim that humans use such a specific rule to pass the false belief task. First, this amount of specification would require a different rule for passing the location transfer task and other versions of the false belief task such as the unexpected contents task. This would mean that evolution had selected for a large set of very specific theory of mind rules, which in turn implies that our ancestors faced a sufficient number of situations very similar to the current psychological tasks. Both of these claims appear extremely unlikely. It would also mean that our theory of mind mechanism (whatever it turns out to be) would be wildly computationally inefficient, which also seems unlikely given the energy expense of such cognitive activities. The findings involving the true belief task (Fabricius et al 2010, Hedger & Fabricius 2011; see below) make it even clearer that the rules we use can't be a large set of specific ones (whether behavioral or mental), since that would entail that we use a specific rule which causes us to *fail* the true belief task. There is simply no way that such an unsuccessful rule by itself would survive the natural selection process. It is much more likely that the failure on true belief tasks is a side-effect (spandrel) of a more general rule (viz. the one which the PAR hypothesis posits) which is useful in a large number of other situations (such as false belief situations).

> AND the person shows liking for an object and is heading vaguely in direction of the object
> THEN the person will go to location A. (p. 5)[75]

Note that Perner's behavioral rule retains some conception of perception ("sees an object") and desire ("shows liking for an object"). Developmental researchers have noted that before children acquire the concepts of desire and perception as mental representations, they use the words 'see,' 'like,' etc. to refer to what we might call behavioral conceptions of perception and desire (Gopnik & Wellman 1994, Perner 1991). Initially, visual perception is conceived of as immediate visual contact with an object of perception (Chapter 3, Flavell 1988), and desire is conceived of as a drive toward some object (Wellman & Woolley 1990). One way of distinguishing these non-representational concepts of proto-perception and proto-desire from their full-blown representational counterparts, and a reason for calling them "behavioral," is that the former are completely transparent from the third person point of view. If a child at this stage of cognitive development notes that S is appropriately oriented with respect to some object, or (perhaps) has her eye gaze directed at some object, then S sees the object, for that's all visual perception is at this stage. Similarly, if S walks across the room and picks up an object, then S wants that object. Desire doesn't require anything going on in a person's head at this stage. Since these non-representational conceptions don't involve the mind or representation, then they might be included in a behavioral rule (see Note 73).

This procedure for collapsing mental rules into behavioral ones can be generalized into Perner (2010)'s "Recipe for posing Povinelli's challenge:"

---

[75] See Note 74 on the specificity of this rule.

> Let the Mentalist give you a specification of the inference procedures required to get from what the animal observes to the ensuing mental state and from that state to the predicted behavior. Collapse these inference rules into one that allows inference of the predicted behavior directly from what the animal observes. (p. 243)[76]

The challenge, then, is to devise an empirical test which could determine whether a given subject is using the *mentalist* or the *behavioral* inference rule(s). In the following two sections I present a way for developmental psychologists to answer this challenge, based on the PAR hypothesis, the previously reported findings of failure on the true belief task (Fabricius et al 2010, see Chapter 2), and the different nature of representational mental states on the one hand and situation-specific behavioral rules on the other.

3. The PAR Hypothesis

Fabricius and colleagues have presented evidence that many 4- and 5-year-olds pass the false belief task by using a behavioral rule which they dub *Perceptual Access Reasoning* or PAR (see Chapter Two; Fabricius & Imbens-Bailey, 2000; Fabricius & Khalil, 2003; Fabricius et al, 2010; Hedger & Fabricius, 2011; Hedger, 2016). PAR involves two general behavioral rules which make reference to behavioral cues and an agent's perceptual access to her immediate environment, but does not involve an understanding of beliefs (as representational mental states). These two rules are:

Perceptual Access Reasoning (PAR):

Rule 1: Seeing → Knowing (and Not Seeing → Not Knowing)

---

[76] Lurz & Krachun (2011) provide several examples of this method applied to various protocols from chimp studies.

Rule 2: Knowing → Getting it Right (and Not Knowing → Getting it Wrong)

Crucially, the use of these two rules cause children to pass the standard false belief task without understanding beliefs.

Thus, as explained in more detail in Chapter Two, using the PAR rules would cause a child to pass the standard 2-location false belief task (Wimmer & Perner, 1983). A child using the PAR rules will reason that since Maxi doesn't see the chocolate, then he doesn't know where it is, and hence Maxi will "get it wrong" and go to the location where the chocolate is not currently located. As Fabricius and Imbens-Bailey (2000) explain, the only incorrect location (in the sense that the chocolate is not there) is the false belief location, so that success on this task is ambiguous between belief attribution and ignorance attribution. Understanding belief would cause the subject to pass, but so would using PAR. The confound in this task can be removed by adding a third, irrelevant, location (Fabricus & Khalil 2003). A child who is attributing a false belief to Maxi should choose the false belief location, but a child using PAR has no reason to choose this location over the irrelevant one, because both locations are equally incorrect. A child using PAR should fail the 3-location false belief task about half of the time.

Fabricius and Khalil (2003) found that, consistent with the PAR hypothesis, many children passed the standard tasks and failed the modified versions by choosing the irrelevant alternatives.[77] The PAR rules are similar to Perner's behavioral rule, then, in that they don't

---

[77] Perner and Horn (2003) tested the PAR hypothesis using a variation of the three-alternative false belief procedure designed to be simpler than that used by Fabricius and Khalil, and they concluded that their findings refuted the hypothesis. However, in Chapter Four I provide five theoretical reasons to remain skeptical about their findings (see also Hedger 2016). There is also empirical evidence, since Gonzales et al (2013) failed to replicate Perner and Horn's findings against the hypothesis.

require an understanding of beliefs or mental representation (more on this in the following paragraphs). However, it is different in that it is much more general, and therefore applicable to a variety of situations. The data show that children using PAR consistently fail three very different versions of the false belief task (Fabricius & Khalil, 2003; see Chapter Two), and it seems much more likely that children are using the general PAR rules rather than a whole series of very specific rules (see Note 74).

Fabricius and colleagues use the words “see” and “know” in formulating the PAR rules because children use these words by 3 years of age (Bartsch & Wellman, 1996). Children using PAR also use this language in their own explanations of their predictions given in psychology experiments (Fabricius et al, 2010). However, children at this stage have not yet mastered the full concepts SEE and KNOW. At this stage of cognitive development, $\text{SEEING}_{\text{PAR}}$ is an immediate visual connection to one’s environment (unobstructed line of gaze), rather than a representational mental state. It is a simple causal connection between objects and agents, rather than a more complicated representational relation between states of the mind and the environment (see Chapter 3; cf. Flavell 1988, Gopnik & Wellman 1994, Perner, 1991). Hence there is no possibility of false perceptions on this understanding. For example, the SEE_{PAR} concept can’t account for the possibility that one is looking at an object but failing to notice it or recognize it as such (Lurz & Krachun, 2011 also make this point).

Likewise, $\text{KNOWING}_{\text{PAR}}$ is also behavioral, correlated with correct or incorrect behavior, but not involving mental representation.[78] That this is so can be seen by the fact that children

---

[78] For more on the concept of knowledge in this PAR stage of ToM development, see Chapter 6.

using PAR fail the true belief task (Fabricius et al, 2010). PAR entails the prediction that children will judge that a protagonist will get it wrong regardless of whether the protagonist has a false belief or a true belief, as long as the situations in both cases result in comparable lack of perceptual access. Thus Fabricius and colleagues (2010) make the novel prediction that children using PAR should fail a true belief version of any of the false belief tasks, such as the unexpected transfer false belief task discussed above. Since either Maxi witnesses the mother move the chocolate (Study 2), or she moves the chocolate back to its initial location (Study 1), when Maxi returns he should have a true belief about where the chocolate is. Therefore children using Belief Reasoning (BR) should pass both versions of this task.

However, for children using PAR, when Maxi leaves his perceptual access to the situation is broken, and they do not attribute a belief to him that persists after he leaves the situation. When Maxi returns, children using PAR see him in a new situation, and when asked where Maxi will look, they should reason that since Maxi can't see the chocolate now, he doesn't know where it is, and thus he will "get it wrong" and look in the empty cupboard (for further details see Chapter Two; Fabricius et al, 2010; Hedger & Fabricius, 2011). In Study 2 of Fabricius et al (2010), to take one example, only 45% of the children (mean age 65 months) who passed the false belief location task passed the true belief version. Thus less than half of the children who passed the false belief task were able to successfully pass the true belief task.

Thus $\text{KNOWING}_{\text{PAR}}$ does not involve mental representation, but is only correlated with correct behavior and perceptual access, as can be seen from the PAR rules (cf. Perner, 1991, for independent evidence of this intermediate understanding of knowledge). Children in this stage (roughly 4 and 5 years of age) have moved beyond the *Reality Reasoning (RR)* of 3-year-olds to something more sophisticated. Perhaps 3-year-olds don't consider anything mental whatsoever when predicting conspecific behavior, but simply reason about how the world really is. At the PAR stage, 4- and 5-year-olds reason about an agent's immediate perceptual access to objects in the environment in order to predict behavior. However, lacking a concept of representational mind, this perceptual access is necessarily linked to the current environmental situation, and does not reside "in a person's head" (and therefore travel with that person in space) nor persist through time across different situations. (Fiebich, 2014, makes the same points when considering the cognitive demands of the three stages in section 5 of her paper.)

This last point warrants emphasizing. As adults, mentalizing, or reasoning about other people based upon our inferences of their mental states, has become second nature to us. Mental thinking has become deeply natural, intuitive, and somewhat implicit by this time. (In Fiebich's terminology, we have become *fluent* in theory of mind; see Chapter 4, Fiebich 2014). On the other hand, the behavioral child's way of thinking is completely alien, and so it is difficult for researchers to get into the PAR child's head. Because it is so foreign, it is difficult for us to think from the perspective of someone who lacks an understanding of the representational mind. Developmental psychologists of the theory-theory bent (e.g. Gopnik & Wellman, 1992; Perner, 1988, 1991; Saxe, Carey, & Kanwisher, 2004) have noted a conceptual change of the sort we are discussing, but missing an awareness of this intermediate stage of development seems to have made them forget what a profoundly different mindset any non-representationalism of this form really comprises. Hence most developmental psychologists have assumed that children were attributing beliefs when they passed the false belief task, perhaps since this is how adults would presumably reason.[79]

PAR reasoning is truly behavioral, in that reasoners at this stage lack a full concept of representational mental states. Thus when an agent witnesses the transfer of an object but then leaves (as in Fabricius et al, 2010, Study 2), PAR reasoners don't attribute a true belief of the object's location to the agent which stays with the agent, but instead reason that when the agent returns, he won't have visual access to the object in this new situation, and hence won't know where it is, but will "get it wrong" and look for the object in the incorrect location. Even the "perception" involved in Perceptual Access Reasoning is not representational. Thus there is no distinction between *seeing* and *looking* for the PAR child.[80]

4. Answering the Challenge

I agree with Perner that any mentalist rule can be collapsed into a behavioral rule by removing the reference to mental states in the mentalist rule. Indeed, in Chapter Two I argue that many children aged 4 and 5 pass the false belief task by using such a behavioral rule, namely the Perceptual Access Reasoning (PAR) rules of Fabricius and colleagues (Fabricius &

---

[79] Thus even other researchers who have found failure on true belief tasks (Friedman et al, 2003; Ruffman, 1996) assume that children are attributing beliefs but using the rule inconsistently. Hence these psychologists failed to recognize that their findings were indicative of a general pattern of behaviorist reasoning and not unique to particular tasks. (See Fabricius et al, 2010, for further discussion of this point.)
Interestingly, we might invoke Chomsky (1965)'s distinction between *competence* and *performance*, but in order to avoid a mistaken inference which is the reverse of the one which worried the renowned linguist. Chomsky was concerned that some might infer from children's language performance errors in some situations that they weren't competent in language grammar, contra Chomsky's nativism. In contrast, many developmental psychologists have mistakenly inferred that children are competent belief reasoners because of their successful performance on false belief tasks, failing to recognize that other factors might explain this situation. In both situations, it's important to remember that competence and performance can come apart.
[80] See Chapter Three for more on perception and the PAR hypothesis, and Chapter Six for more on the concept of knowledge involved in the PAR hypothesis.

Imbens-Bailey, 2000; Fabricius & Khalil, 2003; Fabricius et al, 2010; Hedger & Fabricius, 2011; Hedger, 2016).[81] However, there nonetheless remains a way to empirically test whether a child is using a mentalist or a behavioral rule. The key is that although any mentalist rule can be collapsed into a behavioral one, PAR is a behavioral rule which can't be unpacked into a mentalist one, as the aforementioned failure on the true belief task makes clear.

Because PAR is based upon an agent's perceptual access in the immediate situation, experimental conditions can be devised which cause children using PAR to fail the true belief task. The PAR rule is apparently triggered by a behavioral/situational input involving a temporary absence of the agent and then either the agent's return (Friedman et al, 2003) or some movement of the object in question (Fabricius et al, 2010). Hence the recipe for causing a PAR theorist to fail a true belief task, as outlined in Fabricius et al (2010) and Hedger and Fabricius (2011), involves some cue for *situation change* which causes the subject to reapply the PAR rules to the present situation where perceptual access is lacking. "Situation change" is an intentionally vague parameter, because what constitutes situation change—and hence a triggering mechanism for PAR—is an empirical question to be discovered, not a theoretical question to be answered *a priori* ("from the armchair" as it were). However, if a child understands the concept of a mental representation which persists over time regardless of a momentary break with the agent's immediate environment, then it would simply make no sense to apply the rule again for a new situation whenever the agent returns.

[81] Recall that the concepts SEE and KNOW utilized in the PAR rules are not identical with the fully mental concepts which will eventually be acquired (except perhaps in cases of autism). The words 'see' and 'know' are used by children as early as 3 years of age (Bartsch & Wellman, 1996), but it takes longer for children to exhibit competent facility with the full concepts. As explained in the above sections, SEE refers to immediate perceptual contact between the agent and the situation, and KNOW refers to exhibiting the correct behavior (or "getting it right").

In other words, the very fact that a person fails the true belief task is evidence that the person is using a behavioral rule and lacks an adequate understanding of mental states. An understanding that beliefs and memories of perceptions are representations that persist through time would be sufficient to pass the true belief tasks used in Fabricius et al (2010). Any rule which utilized an understanding of representational mental states should allow one to pass the true belief task. Therefore the true belief task is a sufficient empirical test of whether a given subject is using a behavioral or a mentalist rule to pass the false belief task. As I have explained, a subject using a truly mentalist rule should not fail a true belief task, while those using a behavioral rule such as PAR should fail.[82]

Furthermore, Fabricius et al (2010)'s studies reveal an interesting U-shaped developmental trajectory on the true belief task (see Fig. 1, Chapter Two). Children aged 3 ½ successfully passed the task (83-86% successful in Study 1), while success drops near chance levels for children aged 4 ½ and 5 ½. It's not until 6 ½ years of age that children's success returns to previous levels. This U-shaped development is evidence against the common explanation that ToM tasks might be passed by using behavioral rules which are developed through a process of associative learning (Perner, forthcoming; Vonk & Povinelli, 2006). Here's the general idea. Perhaps rather than using a mental rule to pass the false belief task, children instead use past experiences to learn a contingent statistical regularity to the effect that people typically return to the last place they saw the object (see also the priming explanation of Perner & Ruffman, 2005; more on this below).

---

[82] Thus, the same two-task battery recommended in Chapter Two as a test of ToM ability can also be used to answer the challenge posed by Povinelli and Perner.

However, this hypothesis is not supported by the data concerning true belief tasks. The situation is similar to the well-known phenomenon from linguistics involving children's acquisition of irregular verb past tenses in English. Children are known to correctly use 'went' early in language production, but then go through a period of incorrectly using 'goed' before returning to previous successful past tense production (see Pinker, 1995, for an extensive discussion). The standard explanation for the intermediate period when children say 'goed' is that they are overgeneralizing a rule they have learned (namely, *add '-ed' to form the past tense of a verb in English*). It couldn't be explained by associative learning because the adults they interact with presumably don't speak that way, so there is no statistical regularity to learn.

Similarly, since human agents don't normally return to an incorrect location when they have a true belief about where an object is, it is extremely dubious that children could learn a contingent statistical regularity which would cause them to fail the true belief task (especially considering that they tend to pass at earlier ages). Here as in the case of language acquisition, the most plausible explanation is that children are applying a rule (viz. the PAR behavioral rules) in cases where it doesn't really apply. Thus, the findings of failure on the true belief task provide evidence that a valid empirical test for deciding whether a child or infant is using a genuine understanding of mental state representation in order to pass the false belief task is to include a true belief task.

Incidentally, the true belief task would also be an empirical test for comparing our behavioral explanation of infant success on the false belief task (see Hedger & Fabricius, 2011 for a review) with the one offered by Perner and Ruffman (2005). Perner & Ruffman (2005) argue that infants may be primed to expect constellations of a person, object, and location. For example, in the false belief location task described above, infants may be primed to expect the constellation {Maxi, chocolate, red cupboard}, and will be surprised when Maxi later goes to the green cupboard (recall that the chocolate is currently out of sight). This explanation would predict success on the true belief task, since infants will again be primed to expect the constellation {Maxi, chocolate, red cupboard}, and in this case the true belief location contains the exact same constellation. However, if infants are using some implicit version of PAR (which Hedger and Fabricius call "Rule A"), then we would predict that infants fail a Violation of Expectation (VoE) true belief task (see Chapter Two; Hedger & Fabricius, 2011; and Hedger 2016 for details).

Lastly, what about Perner's worry that we can't know whether even we (adults) are using a theory of mind? Given the ubiquity of folk psychological explanations in our everyday lives, it's a little difficult to take this worry seriously (cf. Fodor, 1987; Doherty, 2009). Although I acknowledge the problems with taking persons' verbal reports of their mental states and processes at face value (Nisbett & Wilson, 1977; Schwitzgebel, 2011; Wilson, 2002), I also think that without evidence to the contrary, we shouldn't think that every person is wildly mistaken. Furthermore, the fact that most people are able to explicate the belief/desire syllogism and apply it with such facility in everyday practice is good evidence that they are indeed using a theory of mind (Fodor, 1987). After all, Perner's behavioral rules are parasitic upon mental rules, as he himself acknowledges. Finally, it would be very surprising if we found that adults did worse than 6 and 7 year olds on a verbal true belief task.

5. Conclusion

The challenge issued by Povinelli and Perner is to devise an empirical method for distinguishing subjects who pass the standard false belief task (and more generally any ToM test) by using mentalist rules from those who pass by using behavioral rules. We first clarified the difference between mentalist and behavioral rules in this context, because by “behavioral rules” Perner (2010, forthcoming) does not mean devoid of intentionality altogether. The main thing which separates mental rules from behavioral ones is that the former require an understanding of mental representation, and the latter lack such an understanding.

Hedger and Fabricius (2011; see Chapter 2) advocate the two task battery of a false and true belief task as a test for theory of mind, and this battery is also able to answer Perner and Povinelli’s challenge. The combination of passing the (standard 2-location) false belief task and failing the true belief task indicates use of behavioral rules, since it exhibits a clear lack of mental representation understanding. Intuitively, the true belief task has always been thought to be easier to pass than the false belief task, which may be one reason why psychologists and philosophers have previously considered it much less important (Hedger & Fabricius 2011). Indeed, most 3-year-olds pass the true belief task (Fabricius et al 2010). The PAR hypothesis is the most plausible explanation of the pre-theoretically unexpected result that 4- and 5-year-olds fail this task yet pass the false belief task. The PAR rules only make sense in their logic and application if the subject lacks an understanding of mental representation. Therefore, the combination of passing the false belief task and failing the true belief task is evidence that subjects are using behavioral rules (viz. perceptual access reasoning) and not mentalist ones.

On the flipside, passing both tasks is evidence that subjects are using mentalist rules and have some understanding of mental representation.[83] Inference to the best explanation leads to this conclusion. I assume this conclusion won't be contentious amongst philosophers and psychologists, since at the present time most of them argue that for humans belief understanding is either innate (e.g. Carruthers 2013, Fodor 1992, Leslie 1987, 1994, 2005, Luo & Baillargeon 2010, Southgate 2014) or is acquired by the age of 4 (e.g. Gopnik & Wellman 1992, 1994, Perner 1988, 1991, Saxe et al 2004, Wellman et al, 2001). Fabricius and colleagues (2010) find that subjects don't pass both tasks until around 6 years of age, and by that age I don't think many cognitive scientists feel that the Perner-Povinelli issue remains a serious concern. Thus, as we have seen, the two task battery of a false and true belief task which Hedger and Fabricius (2011; Chapter Two) advocate as an empirical test for theory of mind is able to answer Perner and Povinelli's challenge.

Povinelli (2001, Vonk & Povinelli 2006) first points out the problem for studying *chimpanzee* theory of mind, and it's interesting to note that there is some evidence that chimps also use Rule A, the behavioral precursor which Hedger and Fabricius (2011) hypothesize is used by human infants and may persist into adulthood as a System 1 cognitive mechanism. For example, chimps in Kaminski et al (2008, Study 2) displayed correct performance on false belief tasks and incorrect performance on comparable true belief tasks (all of this is laid out in more detail in Hedger & Fabricius 2011; see Chapter 2, Chapter 4 and Hedger 2016 for more details concerning the Rule A hypothesis). Therefore the two task ToM battery can also be used as a

[83] Passing the true belief task and failing the false belief task is evidence that subjects are using reality reasoning (RR). See Chapter 2 and Hedger & Fabricius (2011).

valid ToM test for infants and chimpanzees that answers the challenge issued by Povinelli and Perner. Thus the PAR hypothesis is able to solve a current problem in Cognitive Science. It also has further implications for philosophy of mind.

Concepts are the constituents of thoughts. According to the HOT theory of consciousness, in order for a mental state M to be conscious, the subject must have a higher order thought (HOT) *about* that state (Carruthers 2000, Gennaro 2004, Rosenthal 1986). Thus, in order to have conscious beliefs, the subject must have a thought about that belief; perhaps 'I believe that M'. But in order to have such a thought, the subject must have the concept BELIEF. The same is true for conscious perceptions and desires. However, as we have seen, children lack the concept BELIEF until about the age of six. Hence a consequence of the HOT theory would be that children under the age of six cannot have conscious mental states.[84]

This problem with HOT theory, which Van Gulick (2006, p. 13) dubs the "too fancy" objection, has been discussed previously (Dretske 1995, Tye 1995). However, now the objection can be made more precisely, and there is clear empirical evidence in favor of it. Now we can say exactly what it is that is too sophisticated for a young child to possess, viz. an understanding of mental representation. Furthermore, the objection is also considerably strengthened. According to the PAR hypothesis, children don't acquire such an understanding until they reach the third developmental stage of Belief Reasoning (BR). This doesn't happen until around six years of age. Previously, the "too fancy" objection was thought by many to apply only to

[84] Similar reasoning could expand this line of argument to apply to animals as well, and indeed the problem is often posed in those terms. Although I won't take the time to do it here, I think this argument can also be adapted fairly easily to apply to self-representational theories of consciousness (Kriegel 2003, 2009, Kriegel & Williford 2006). See also Chapter 7, Section 3.3 for more on this topic.

infants, and not to school age children (cf. e.g. Gennaro 2012). Although I find the idea that infants don't have phenomenal experience highly implausible, there have been philosophers who have notoriously bitten the bullet on this point (Carruthers 2000, 2005). However, it's even less plausible to deny that 5-year-olds have conscious mental states.

In order for an experience to be phenomenally conscious, at a minimum the subject must be explicitly aware of that experience. Since young children don't understand mental representation, it follows from the PAR hypothesis that they can't experience beliefs and perceptions *as representations*, even for those who don't subscribe to HOT theory and don't deny that they have conscious mental states. Hence another implication of the PAR hypothesis is that the phenomenal experience of young children might be importantly different from that of adults, at least insofar as adults (at least sometimes) are aware of e.g. our visual experiences *as representations*. This might occur when we experience visual illusions such as the Müller-Lyre Illusion or a stick bent in water (to take two of philosophers' favorite examples). Even common experiences such as mistaking things in the distance for people, or wondering whether what we see in our backyard in the evening is an animal or something else, would seem to involve understanding that visual perceptions are representations and can therefore sometimes *mis*represent. Thus it would seem these kinds of experiences would be unavailable to children younger than six.[85]

[85] This line of thought is further developed in Chapter 7, Section 4.

References

Bartsch, K., & Wellman, H. (1996). *Children talk about the mind.* New York, NY: Oxford University Press.

Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory.* Cambridge, England: Cambridge University Press.

Carruthers, P. (2005). *Consciousness: Essays from a higher-order perspective.* Oxford, England: Clarendon Press.

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

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: The MIT Press.

Doherty, M. J. (2009). *Theory of mind: How children understand others' thoughts and feelings.* New York: Psychology Press.

Dretske, F. (1995). *Naturalizing the mind.* Cambridge, MA: MIT Press.

Fabricius, W. V., Boyer, T., Weimer, A.A., & Carroll, K. (2010). True or False: Do five-year-olds understand belief? *Developmental Psychology*, *46*, 1402-1416.

Fabricius, W.V., & Imbens-Bailey, A. L. (2000). False beliefs about false beliefs. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning about the mind* (pp. 267-280). Hove, England: Psychology Press.

Fabricius, W. V., & Khalil, S. L. (2003). False beliefs or false positives? Limits on children's understanding of mental representation. *Journal of Cognition and Development, 4*, 239-262.

Fiebich, A. (2014). Mindreading with ease? Fluency and belief reasoning in 4- to 5-year-olds. *Synthese, 191,* 929-944.

Flavell, J. H. (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J.W. Astington, P.L. Harris, & D.R. Olson (Eds.), *Developing theories of mind*. New York: Cambridge University Press.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind.* Cambridge, MA: MIT Press.

Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition, 44,* 283-296.

Friedman, O., Griffen, R., Brownell, H., & Winner, E. (2003). Problems with the seeing = knowing rule. *Developmental Science, 6*, 505–513.

Gennaro, R. J., Ed. (2004). *Higher-order theories of consciousness: An anthology.* John Benjamins.

Gennaro, R. J. (2012). *The consciousness paradox: Consciousness, concepts, and higher order thoughts.* Cambridge, MA: MIT Press.

Gonzales, C., & Fabricius, W. V. (2013). Four-year-olds' performance on three-alternative false belief tasks: Evidence for perceptual access reasoning. Poster presented at Cognitive Development Society VIII Biennial Meeting, Memphis, 2013.

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7,* 145-171.

Gopnik, A., & Wellman, H. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture*, pp. 257-293. Cambridge, England: Cambridge University Press.

Hedger, J. A. (2016). Perceptual access reasoning: Developmental stage or System 1 heuristic? *Phenomenology and the Cognitive Sciences, 15,* 207-226.

Hedger, J. A., & Fabricius, W.V. (2011). True belief belies false belief: Recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development. *Review of Philosophy and Psychology, 2,* 429-447.

Kriegel, U. (2003). Consciousness as intransitive self-consciousness: Two views and an argument. *Canadian Journal of Philosophy, 33,* pp. 103-132.

Kriegel, U. (2009). *Subjective consciousness: A self-representational theory.* Oxford: Oxford University Press.

Kriegel, U., & Williford, K., Eds. (2006) *Self-representational approaches to consciousness.* Cambridge, MA: MIT Press.

Leslie, A. M. (1987). Pretense and representation: The origins of 'theory of mind.' *Psychological Review, 94,* 412-426.

Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In *Mapping the mind: Domain specificity in Cognition and Culture,* Ed. L. Hirschfeld and S. Gelman. Cambridge, England: Cambridge University Press.

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

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

Lurz, R. W., & Krachun, C. (2011). How could we know whether non-human primates understand others' internal goals and intentions? Solving Povinelli's problem. *Review of Philosophy and Psychology, 2,* 449-481.

Lycan, W. G. (1996). *Consciousness and experience.* Cambridge, MA: MIT Press.

Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

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

Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind." *Philosophical Transactions of the Royal Society B—Biological Sciences, 362,* 731-744.

Perner, J. (1988). Developing semantics for theories of mind: From propositional attitudes to mental representation. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 141-172). New York: Cambridge University Press.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: The MIT Press.

Perner, J. (2010). Who took the 'cog' out of cog*nitive* science? Mentalism in an era of non-cognitivism. In P. A. Frensch & R. Schwarzer (Eds.), *Cognition and neuropsychology: International perspectives on psychological science, Volume 1,* pp. 241-261. ??,??: Psychology Press.

Perner, J. (forthcoming). Theory of mind—an unintelligent design: From behavior to teleology and perspective. In A. M. Leslie & T. C. German (Eds.), *Handbook of 'Theory of Mind'.* Piscataway, NJ: Lawrence Erlbaum Associates.

Perner, J., & Horn, R. (2003). Knowledge or false negatives: Do children of 4 to 5 years simulate belief with "not knowing = getting it wrong?" *Journal of Cognition and Development, 4*, 263-273.

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

Pinker, S. (1995). *The language instinct: How the mind creates language.* New York: HarperCollins.

Povinelli, D.J. (2001). On the possibilities of detecting intentions prior to understanding them. In B. Malle, D. Baldwin, & L. Moses (eds.), *Intentions and Intentionality: Foundations of Social Cognition.* Pp. 225-248. Cambridge, MA: MIT Press.

Povinelli, D.J. & Vonk, J. (2003). Chimpanzee minds: Suspiciously human? *Trends in Cognitive Science, 7*, 157-160.

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

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies, 49,* 329-359.

Ruffman, T. (1996). Do children understand the mind by means of simulation or a theory? Evidence from their understanding of inference. *Mind and Language, 11*, 388 -414.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55,* 87-124.

Schwitzgebel, E. (2011). *Perplexities of consciousness.* Cambridge, MA: MIT Press.

Southgate, V. (2014). Early manifestations of mind reading. In S. Baron-Cohen, H. Tager-Flusberg & M. Lombardo (Eds.), *Understanding Other Minds*, 3rd Edition. Oxford University Press.

Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind.* Cambridge, MA: MIT Press.

Van Gulick, R. (2006). Mirror mirror—Is that all? In Kriegel & Williford (2006), pp. 11-39.

Vonk, J. & Povinelli, D.J. (2006). Similarity and difference in the conceptual systems of primates: The unobservability hypothesis. In E. Wasserman and T. Zentall (Eds.) *Comparative Cognition: Experimental Explorations of Animal Intelligence.* pp. 363-387. Oxford University Press.

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.

Wellman, H. M., Lane, J. D., LaBounty, J., & Olson, S. L. (2011). Observant, Nonaggressive Temperament Predicts Theory of Mind Development. *Developmental Science*, *14*(2), 319–326.

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35,* 245-275.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: Belknap Press.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-28.

CHAPTER SIX

Explaining Cognitive Development as a Change in Conceptual Repertoire:

Defending a Limited Theory-Theory of Concepts against Fodor's Shareability Concern

"Acceptance of some form of intentional explanation appears to be a cultural universal. There is, so far as I know, no human group that doesn't explain behavior by imputing beliefs and desires to the behavior. (And if an anthropologist claimed to have found such a group, I wouldn't believe him.)" –Jerry Fodor[86]

"Experiments like this are important because they show us that even though children use the same words we use, they may not mean the same things by them." –Janet Wilde Astington[87]

1. Introduction

There is a move which has become quite familiar in developmental psychology, which arguably goes back to Piaget and the very beginnings of the field. Children of various ages are given certain experimental tasks, and researchers find that a significant number of those children frequently give answers which are not only wrong, but are—at least from a normal adult point of view—downright bizarre. This is seen by researchers to be best explained by the fact that the child's thinking about the world must be deeply and fundamentally different from that of the adult. From this, many researchers conclude that some children must either lack concepts that most adults share, or else have only an incomplete or partial concept which the

[86] Fodor (1987), p. 132.
[87] Astington (1993), p. 130.

adult grasps fully.[88] Very often, as with the case of Piaget, development is then seen to progress through certain *stages*, which can be defined by the concepts and understandings that children have at different times during their cognitive development.

Take as an initial example Piaget's well known *conservation task*. This can take many different forms, but a typical task is the following. First, children are shown two identical glass containers, each holding an equal amount of liquid, and are asked if the two amounts of liquid are the same. Of course, most children successfully answer "Yes." While the child observes, the psychologist takes the liquid contents of one of the glasses and pours it into a taller and narrower glass. Piaget and other researchers found that many children aged 2 to 6 years will now say that the two amounts of liquid are *different*, because the thin glass contains *more* liquid than the original glass contained.[89]

For the normal adult, this sort of mistake is quite surprising. Of course an adult who witnessed the same amount of bourbon in 2 different types of glasses, served at a tavern for instance, might mistakenly think that her friend has received a larger portion of bourbon than she herself has. However, an adult who witnessed the bourbon being poured from one glass to another would never think that the amount of bourbon has somehow changed simply as a result of the act of pouring it into a new glass, regardless of the phenomenal appearance of either glass of bourbon. Hence, children at this stage of development (which Piaget dubbed

---

[88] Part of the reason for thinking children have different *concepts* is that they (1) use the same words that adults do, but at times seem to mean something very different by them; and (2) this difference in meaning is not easily explained by a simple misapplication of the word to a different, shared concept (as is the case in the well-known phenomenon of *overextension* when a child acquiring English may use the word 'cat' to mean all animals e.g.).

[89] See Piaget & Inhelder (1974) for more on these highly replicated studies.

"pre-operational") must be thinking about the world in a radically different way, in order to even think that such a thing is possible. Specifically, children at this stage don't seem to realize that there are some properties of a thing which are invariant and persist through many types of physical transformation. (This was further corroborated by the fact that children made similar mistakes in other cases, such as when a spherical ball of clay is rolled into a snake-like shape, etc.)[90]

Although Piaget's theory of defining stages of cognitive development in terms of different logical systems has been relegated to the dustbin,[91] the intuitive plausibility of explaining the behavior of children in terms of a unique conceptualization of the world remains. This paper will attempt to sketch such a story for the children's concept KNOW at an intermediate stage of theory of mind (ToM) development, with the assumption that a similar story could be generalized to a limited number of other cases. This story will draw on recent empirical developments in the theory of mind, and in particular the Perceptual Access Reasoning (PAR) hypothesis of William Fabricius and colleagues (see Chapter 2). Considering concerns about theory-theory as a philosophy of concepts (defended by Carey 1985, 2010; Gopnik, 1988; and others), a defense will be sketched to one of Jerry Fodor's major criticisms

---

[90] After writing this introduction I came across Carey's (2010) description of the same phenomenon, which she labels the "Huh?" factor (pp. 357-358).

[91] That this is so is evidenced by Perner's (1991) characterization of a large amount of the developmental psychology of the 1970's and 1980's as "Piaget Bashing" (see Doherty 2009, pp. 12-13, for a succinct discussion). The reasons for this are largely twofold: Empirically, there are a vast number of studies demonstrating that children of a given age, which are not supposed to be able to do such-and-such according to Piaget's theory, can actually do such-and-such given appropriate task adjustments (see Gelman & Baillargeon 1983, and Perner 1991, for partial reviews). That Piaget's theory is so general is both a virtue and a vice in this regard. Second is the more philosophical point that it doesn't seem right to call most children's errors *logical* errors, since in many cases they're making valid inferences given their limited understanding (see e.g. Carey 1985, p. 75). (Although they didn't have as much impact in the cognitive science community, one might also add Fodor 1975's argument that Piaget's logics aren't the sort of things that can be learned.)

for such proposals, which Kevan Edwards (2010) calls the "Shareability Constraint."[92] First, however, a momentary digression on theory-theories in general which I hope will strengthen the case for the theory-theory of concepts, in addition to clarifying what has been construed by some as a confusing terminology.

2. Digression on Theory-Theories: A Happy Marriage of Distinct Views

Many, including Fodor (1998, p. 112), Laurence & Margolis (1999, p. 43), and even Carey herself (2010, p. 484) have complained that calling the Susan Carey (1985, 2010) view the "theory-theory" of concepts is unfortunate, since it's too easily confused with the "theory-theory" of mental state attribution (see e.g. Gopnik & Wellman, 1992, 1994). In this section I will sketch three different versions of a *theory-theory* view in psychology, and show that although these views are logically distinct and could therefore be held in isolation from each other, they nonetheless form a happy tripartite union. When one holds all three views together, they mutually reinforce one another in a way that makes each of them more plausible as an inference to the best explanation for the observed psychological data. Or so I shall argue. In what follows, reviewing all of the relevant experimental data in support of each view would take a book-length treatment, so we will have to make do with one representative key finding for each view. To get a fuller picture, one should read the cited papers and books.

2.1 The Theory-Theory of Mindreading

[92] Perhaps I should flag here that I've changed the spelling from Edwards (2010), who attributes the terminology to Georges Rey. Fodor (1998 and elsewhere) calls it the "publicity constraint."

The theory-theory view of mental state attribution (Gopnik & Wellman, 1992, 1994) contends that children utilize folk psychological theories in order to attribute mental states both to themselves and others, and also to use these attributions to predict behavior. According to the PAR view (Chapter 2, Fabricius et al 2010, Fabricius & Imbens-Bailey 2000, Fabricius & Khalil 2003, and Hedger & Fabricius 2011), there are three distinct stages of cognitive development, each of which makes use of a different theory.[93] The Reality Reasoning (RR) stage (roughly age 3) is associated with the theory that there is a direct (i.e. unmediated) causal connection from reality to behavior. Children in this stage tend to predict that agents always go to the correct location in the false belief task, and lack the understanding that the same thing can appear differently to different people, depending upon their visual perspective (Flavell 1974's "Level 2" perspective taking).[94] The Perceptual Access Reasoning (PAR) stage (roughly age 4 to 5)[95] applies a series of two rules which take as input a representation of an agent with a goal, initiate an analysis of the agent's perceptual access to the goal (e.g. food) in the present situation, and output a prediction of that agent's behavior (see Chapter 2, Fabricius et al 2010, Fabricius & Imbens-Bailey 2000, Fabricius & Khalil 2003, and Hedger & Fabricius 2011 for more on the PAR stage). These first two stages do not use any *representational* mental state concepts; instead, they are purely causal/behavioral (more on this below). Finally, the

[93] This is a modification of the Gopnik & Wellman (1992, 1994) view, which contains only two stages and a shorter timeline, but otherwise the theory-theory aspects of the two views remain essentially the same.
[94] For those who are unfamiliar with these experiments, see Chapter 2 or Hedger & Fabricius (2011) p. 430 for a description of the false belief location task, and Chapter 3 or Masangkay et al (1974) for examples of Level 1 and Level 2 perception tasks.
[95] PAR is used in explicit reasoning and verbal reports in children ages 4 and 5, but a version of PAR (dubbed "Rule A") may be implicitly used as early as 1 year of age, and may persist into adulthood. See Chapters 2, 4 and 5 and Hedger 2016 for the details of this hypothesis, and Hedger & Fabricius (2011) for a review of infant and chimpanzee theory of mind studies from the perspective of the PAR view.

Belief Reasoning (BR) stage (roughly age 6 ½ to adult for psycho-typical subjects) utilizes a full-blown theory of mind, in the sense that it makes use of a (perhaps quite complicated) folk psychological theory which predicts behavior based upon mental states such as beliefs and desires.

According to theory-theory, at each stage subjects apply a theory to predict behavior and (in the case of BR, at least) attribute mental states to themselves and others. One example of a finding which supports this view is that children do not give the correct answer on a task about someone else's false belief until they are able to give the correct answer on a task concerning their own false belief. For instance, in the unexpected contents false belief task, a child is shown a familiar candy container, such as an M&M bag, and asked what he thinks is inside. After the child says "M&M's," it is revealed that something unexpected is inside, such as a pencil. After the pencil is placed back inside the bag, the subject is informed that a friend of the experimenter's (named "Elmo" e.g.) is waiting outside. The experimenter then asks, "If he just looks at it, what will Elmo think is inside the bag?"

Wimmer & Hartl (1991) and others found that 3 year olds (RR stage) typically incorrectly infer that someone else will believe that the candy bag contains a pencil. Surprisingly, 3-year-olds also report that they initially believed a pencil was in the bag, even though they just a minute ago reported that they believed that the bag contained candies. Further experiments were done in order to rule out conflicting explanations, such as that the children didn't remember or were embarrassed about giving an incorrect response. This experiment has been replicated several times. Interestingly, children don't correctly attribute a false belief about themselves until they are able to do so for others (Astington & Gopnik 1988, Gopnik & Astington 1988, Wimmer & Hartl 1991). This is exactly what would be expected if subjects were applying a theory in both the first-person and third-person cases, and is contrary to the common sense Cartesian view of mind that we have a privileged epistemic access to our own mental states which we do not have when it comes to the mental states of others.[96] According to the latter sort of view, first-person attribution should be decidedly easier than third-person attribution. However, that doesn't appear to be the case.

2.2 The Nisbett and Wilson View

There is also a view about our introspective access to our own mental states from social psychology which might aptly be called *theory-theory (*Nisbett & Wilson, 1977; Wilson, 2002). These researchers found that subjects consistently give reports of their own mental states which are demonstrably incorrect, or may behave as though they are self-attributing a mental state which they probably don't have (Nisbett & Wilson, 1977; Wilson, 2002). For instance, in one study which is typical, a female experimenter (herself blind to the experimental hypothesis) stopped men and asked them to fill out a questionnaire (Dutton & Aron 1974). She then gave the subjects her phone number and invited the men to call if they were interested in obtaining the survey results. There were two conditions: in one she stopped them on a safe bridge (10 ft. high) and in the other she stopped them while crossing a more precarious feeling suspension bridge 230 ft. high.

---

[96] The Cartesian or privileged access view is defended in some form or other by philosophers such as Andre Gallois (1996), Brie Gertler (2011), Alvin Goldman (2006), and Sydney Shoemaker (1988).

Men were much more likely to call if they had been stopped on the precarious bridge, presumably because they attributed their physiological responses—which were actually caused by the fear of the bridge—to arousal caused by the presence of the experimenter.[97] According to this theory-theory, the data which a subject has introspective access to, such as the physiological responses noted above, often greatly underspecify the mental state that one is in (think e.g. of the physiology of fear, anger, sexual arousal, and surprise, just to name a few). The subject then applies a folk theory about mental states, noting salient aspects of one's environment which may or may not be in fact relevant to one's occurrent state (such as the presence of a female), in order to self-attribute mental states, preferences, and the like (Dutton & Aron, 1974; Nisbett & Wilson, 1977; Wilson, 2002).[98]

Interestingly, the view developed by social psychologists Richard Nisbett, Timothy Wilson, and others is strikingly similar to a view put forth by neuroscientist Michael Gazzaniga (1985), who arrived at the same conclusion from considerations of completely different data. Studying subjects who had the corpus callosum connecting the two brain hemispheres

---

[97] Specifically, 61% of the experimental group members called, whereas only 23% of the control group members called, across two studies. In the first study, the experimental group crossed the "dangerous bridge" whereas the control group crossed the "safe bridge," and in Study 2 both groups crossed the suspension bridge, but the control group was interviewed 10 minutes after crossing.
The theory that environmental cues are used to determine emotion judgments during ambiguous states of arousal was originally put forward in Schachter & Singer (1962).

[98] Bob Van Gulick worries about this particular example, since subjects in the Dutton & Aron (1972) paradigm don't explicitly give explanations of their behavior. However, subjects do give explicit explanations of their behavior in many other studies. For instance, in probably the best known of Nisbett & Wilson (1977)'s experiments, conducted in various stores under the guise of a consumer survey, passers-by were asked which pair they preferred, from a choice of five identical pairs of stockings. Although a bias to choose the right-most object was discovered (by almost four to one), subjects never reported an awareness of this bias and once again confabulated explanations for their preference ("It's the softest," "The material is thicker," "It looks to be of higher quality," etc.) When informed of the position bias effect, subjects denied it outright, giving the experimenters "worried glances."

surgically severed in order to treat extreme epilepsy (so called “split brain” patients), Gazzaniga and others found that behaviors initiated by their right brain hemisphere were often unavailable for verbal report by the left hemisphere. When queried about these behaviors, the left brain (presumably ignorant of the true motives for the actions, which were interpretable based upon the experimental design) often *confabulate* (Gazzaniga’s term) a rational but completely false explanation for them. Furthermore, the subjects are completely convinced of the accuracy of these confabulations.

In one well-known experiment, for example, a picture of a chicken claw is flashed to the right visual field (and then processed by the left brain hemisphere), and a picture of a snow scene is flashed to the left visual field (Gazzaniga & LeDoux 1978, p. 148). The subject was then asked to choose from an array of pictures presented to him. With his left hand (controlled by the right hemisphere) he selected a shovel and with his right hand he selected a chicken. When asked why he chose those items, the subject replied, “Oh, that’s simple. The chicken claw goes with the chicken, and you need a shovel to clean out the chicken shed.” In other words, the left hemisphere fabricates an explanation for the subject’s behavior which is consistent with the knowledge of the left hemisphere but fails to take into account the knowledge of the right hemisphere. Based on over 50 years of studying the brain, Gazzaniga (1985)’s hypothesis is that mental modules[99] operating beneath the level of consciousness are responsible for our behavior, but that a left brain *interpreter module* theorizes a post hoc rationalization of that

[99] See Fodor (1983) for more on the notion of modular mental architecture.

behavior. The interpreter uses a theory that says our behavior is the result of our own free will and is wholly rational.

The connection between the two theory-theory views so far discussed (mentalizing theory-theory and social psychology/neuroscience theory-theory) is, I hope, obvious by now. Both views argue that self-attribution of mental states (including the intentions behind many of our actions) is achieved by applying a theory and post hoc rationalization, rather than by some Cartesian process of introspection. The fact that data from three different fields—social psychology, neuroscience, and cognitive development in children—led researchers independently to the same theory strikes me as quite compelling, and makes a strong case for the truth of the view. Also, it's obvious that the data in support of any of these theories would mutually support the other two, since the theories are essentially the same.

2.3 The Theory-Theory of Concepts

Finally, the theory-theory of concepts (Carey 1985, 2010; Gopnik, 1988) argues that some concepts are (at least partially) individuated by larger theories in which they play a crucial role.[100] Sometimes, and for a very limited number of concepts, different theories give rise to different, incommensurable, concepts. These concepts are incommensurable because they can't be easily carried over from one theory to another. Utilizing theory 2 makes some concepts which were crucial in theory 1 almost incomprehensible. This happens when one theory can't

[100] A caveat which may be important for some is that there doesn't seem to be (as of yet) a fully worked out theory-theory. Fodor, for instance, complains that "it's [often] so hard to figure out what… the theory-theory is a theory of" (1998, p. 117), and Carey (2010, p. 484) concurs that "The theory-theory of conceptual development is not actually a theory [because] it's way too underspecified." The skeleton given here, then, is appropriately sketchy.

be easily translated into the language of the other without some translator's gloss (Carey 1991, 2010). For instance, the adult theory of mind, in which the idea of *mental representation* is central, is radically different from the BR or PAR stage theories of mind, where the idea of representational mental states is completely absent. Things which seem quite natural to us can't even be considered by children at this age.

Without the idea of representation, a whole host of ideas and distinctions—veridical versus illusory, for example—are not possible. The idea of visual illusion, where we have a perception which misrepresents reality, cannot be understood by someone who lacks the notion that mental states represent. Thus Flavell and colleagues have demonstrated that children in the RR stage don't make what they dub the *appearance/reality distinction* (Flavell et al 1983). One task they used to test this involved showing children a sponge which looks like a rock, and then allowing them to handle it. After the child is familiarized with the fake rock, it is placed on the table. Children are then asked, "Is this really a rock or really a sponge?" and "When you look at this with your eyes right now, does it look like a rock or look like a sponge?" They found that before four years of age children tend to say that it looks like a sponge and that it really is a sponge, presumably because they don't realize that things can be different than they appear to be. They don't view perceptions as representations of what is perceived, but more like direct causal links with the environment.

The connection between the theory-theory of concepts and the previous two theory-theories becomes clearer when we focus our attention on mental concepts. Consider the transition from the PAR stage to the Belief Reasoning (BR) stage of theory of mind development (see Chapter 2, Chapter 4, Hedger 2016, Hedger & Fabricius 2011). PAR makes no use of the notion of mental representation, as is aptly demonstrated by failure on the *true belief task* during this stage of development (Chapter 2, Fabricius et al 2010). Fabricius et al (2010) found that although four- and five-year-olds pass the false belief task (which researchers have been using to study theory of mind in children for 30 years), they fail true belief versions of the task. In one such task, children were shown a story involving props and a puppet named Maxi. Maxi places his chocolate in a red cupboard, and then watches as his mother comes in and moves the chocolate from the red cupboard to the green cupboard. Maxi goes outside to play, and there is no movement of the chocolate during his absence. After his return, the child is asked, "Where will Maxi look for his chocolate?"[101] The success rate of 5-year-olds was not significantly greater than chance.

This finding is significant, because all it would take to pass this true belief task is the understanding that Maxi has a representation (belief) of the location of the chocolate which will persist through a brief departure and subsequent return. It's initially quite surprising that anyone would fail such a task, and children at an earlier stage of cognitive development (RR) pass this task although they fail the false belief version. When combined with other findings in theory of mind development, this makes a compelling case that children younger than 6 years of age do not understand mental representation (see the rest of this monograph, especially Chapter 2 and Chapter 5). Also, children in the PAR stage have concepts of mental states such as KNOW which don't map neatly onto the adult concept. Whereas the idea of mental

[101] A meta-analysis of 178 false belief studies found no significant difference between asking "Where will Maxi look?" "Where does Maxi think it is?", etc. (Wellman, Cross & Watson, 2001).

representation is a crucial part of the adult concept KNOW, it is completely missing from the PAR concept.

The theory-theory of mindreading commonly holds not only the thesis that subjects use a theory to attribute mental states to ourselves and others, but also that the mental state concepts of children—concepts such as WANT (desire), SEE (perception), etc.—are partially constituted by the theories which contain them (e.g. Gopnik & Wellman, 1994). If the theory-theory of concepts is right that theory change may include conceptual change (Gopnik, 1985, 2010), and the theory-theory of mindreading is right that mental states are attributed by a process of theory application, then it's natural to think that the theories children use for the latter purpose will have different concepts of mental states (the key theoretical posits) at different stages of cognitive development. When theories that children use define stages of development, as they do in the PAR theory, then it's a short step to thinking that the theories play these two cognitive roles. In this way, all three theory-theories fit naturally together and mutually support each other in one coherent picture. Although logically distinct, once in the abductive frame of mind of a scientist, seeing that the evidence for one theory also supports the other two in a tidy overall scheme leads one to accept them as a single package.

3. The Child's Concept KNOW

According to the PAR hypothesis (Chapter 2, Fabricius et al 2010, Fabricius & Imbens-Bailey 2000, Fabricius & Khalil 2003, Hedger 2016, Hedger & Fabricius 2011), at a certain stage of cognitive development, children demonstrate a complete lack of understanding of mental representation. This is evidenced by their failure on true belief tasks, as discussed in the previous section (see Chapter 2, Chapter 5, Fabricius et al 2010, and Hedger & Fabricius 2011 for more on this). The concept KNOW which children have at this stage of development is importantly different from the adult concept. One way to characterize this difference is to say that the adult concept is a mentalist one while the child's concept is behavioral (see Chapter 5). Since children at the PAR stage lack any understanding of mental representation, their concept KNOW is inextricably linked to behavioral cues and cues about relations in the present situation.

3.1 Knowledge and the PAR Rules

More specifically, PAR theory hypothesizes that during this intermediate stage of theory of mind development, there is a mental mechanism which takes as input an agent, a goal (such as some object of interest), and some perceptual link between them.[102] This perceptual link is not the adult conception that perceptions represent the physical environment, but rather a sort of immediate causal contact in the present situation. Hence, when Maxi leaves and returns in one version of the true belief task, the child considers this perceptual link to be broken. Call this perceptual link "seeing." From this and similar data gathered from the various versions of the *false belief task* (see Chapter 2 and Hedger & Fabricius 2011 for the PAR interpretation of the false belief task results) and Pillow (1989)'s work on the early link between perception and

[102] Of course, "agent," "goal," "object of interest" and the like are already intentional concepts, so even PAR is not pure behaviorism. (Again, the key difference separating PAR from BR is not the absence of mentalism *per se* but the idea that mental states are representations.)
It's not important for present purposes, but another cue to the PAR mechanism is *situation change*. This parameter is left intentionally vague in the theory, because sorting out what constitutes situation change—and hence a triggering mechanism for PAR—is an empirical question. See Hedger & Fabricius (2011) for discussion.

"knowledge", etc., PAR theorists posit the first rule of Perceptual Access Reasoning (Fabricius & Khalil 2003):

Rule 1: Seeing → Knowing (and Not Seeing → Not Knowing)

The reason for the scare quotes around 'knowledge' and 'seeing' in the previous paragraph is that the PAR concepts of perception and knowledge are to be distinguished from the adult concepts. Just as SEE_{PAR} (the PAR stage concept of sense perception) is a non-representational link having to do with positions of agents and goals in the current situation, $KNOW_{PAR}$ is not a representational state either. Rather, knowledge for the PAR child is simply something which predisposes an agent to behave appropriately given her goals, which we might call "getting it right" (see Perner 1991 and below for evidence of this which is completely independent of the PAR theory). Hence, in the location tasks (both false and true belief versions), "knowing" is a label for the predisposition to go to the correct location—i.e., the location where the goal is actually located. In the true belief task described earlier, the PAR child first determines whether the protagonist can "see" the goal in the present situation. If Maxi can "see" the goal, then he "knows;" if he can't "see" the goal, then he doesn't "know."

This is a direct causal link which does not result in a representation in Maxi's mind, so thus when Maxi leaves and returns and the PAR perceptual link is broken, the knowledge does not persist but also becomes a broken link.[103] There is no understanding included in $KNOW_{PAR}$

[103] What is required to break this link is largely an empirical question, not a theoretical one. Hedger and Fabricius (2011) explain that once perceptual access is broken, the child will now deem the current state of affairs as a new situation, and will then re-apply the PAR rules. Discovering these cues for situation change is a job for empirical discovery. Because SEE_{PAR} is a completely foreign concept from the adult point of view (after all, that's part of the point of this chapter), there is no obvious a priori method for discovering these cues. We are able to specify two cues for situation change in the false and true belief paradigms which are evident from published studies. One is to

that people can get things correct by guessing, or that people occasionally make mistakes even if they know. Thus, since Maxi lacks the perceptual connection in the true belief task, then he doesn't know. Since he doesn't know, and there is no getting it right by accident or inference for the PAR child, then Maxi will definitely get it wrong, and look in the location where the chocolate is not currently located. The second rule of PAR is this simple connection between knowledge and appropriate behavior (Fabricius & Khalil 2003):

Rule 2: Knowing → Getting It Right (and Not Knowing → Getting It Wrong)

Hence, PAR Theory therefore provides a definition of the KNOW concept at this stage of development:[104]

$KNOW_{PAR}$ df: S knows iff S demonstrates correct behavior and S has perceptual access.[105]

---

have some movement of the object being sought which is occluded from the agent (Fabricius et al 2010, Kaminski et al 2008), and the other is to have the agent momentarily leave the situation and return (Fabricius et al 2010, Friedman et al 2003). Whether there are other cues which trigger the PAR rules to be computed again, and what those might turn out to be, are matters for empirical investigation (see Chapter 2, Section 4 for more discussion).

[104] As with all rules of this sort, there arises the question whether the two PAR rules are represented somewhere in the child's mind, or simply descriptive reconstructions of mental mechanisms posited by psychologists, which may or may not be explicitly represented anywhere in the child's mind. (This is perhaps related to the Dennett/Fodor dispute over rules and representation, such as Dennett's example of the chess program which is described as "getting its Queen out early." See e.g. Dennett 1978; Fodor 1985, 1987.) In this case, the rules are explicitly and consciously followed by children at this stage, as demonstrated by the fact that many PAR subjects will cite the rules as explanations for their predictions given during false and true belief tasks (Fabricius et al 2010).

[105] It might be thought that rather than the actual demonstration of correct behavior, what is meant here is merely the *disposition* to behave correctly (so that 'demonstrates,' 'exhibits,' etc. in the descriptions here and below should be replaced with modals such as '*would* demonstrate, given appropriate circumstances,' etc.). However, it's not at all clear from the empirical data that children do think of knowledge as a mere disposition. Because it's more of a direct behavioral relation (not mediated mentally), I think it's a better description (and more in line with what children actually say) to refer to the actual demonstration of correct behavior.

This definition, providing both necessary and a sufficient conditions for knowledge, follows straightforwardly from the two PAR rules above.[106] Knowing for the PAR child *just is* the exhibition of appropriate behavior. If S knows, then S will get it right. If S gets it right, then S knows. Similarly, there can be no knowledge without perceptual access.[107] Thus, the simplicity of this concept of proto-knowledge allows us to define it in a way which is impossible for most ordinary, non-technical concepts (Fodor et al 1980). The other notable exceptions to this generalization are technical terms, such as those often employed by philosophers for theoretical purposes ("terms of art"). This lends further evidence to the idea that concepts such as $KNOW_{PAR}$ exhibit a special nature, best explained by the fact that they are partly constituted by the theories in which they play a central role. The hypothesis I wish to endorse is that all mental state terms are like this.

Furthermore, it becomes obvious why the PAR concepts of perception and knowledge will be replaced by the superior representational concepts once those are available (after all, the former lead to the wrong prediction when someone has a true belief). This is perhaps the key to the transition from PAR stage to the BR stage. The nature of this process of concept

---

[106] The first version of each rule can be interpreted as a material conditional of the form 'If A then B.' If the second version of each rule (given in parentheses above) is interpreted as the inverse of the first rule (i.e. 'If not-A then not-B'), then according to classical logic we can use the contraposition of that rule (i.e. 'If B then A') and combine it with the first rule to obtain the biconditional 'A if and only if B.' Hence a definition with necessary and sufficient conditions logically follows from the description of the two knowledge rules in PAR (at least on one possible interpretation).

[107] This is perhaps best demonstrated by the no belief task (see Chapter 2, Section 5). In one version of this task (Chen et al 2015), Lee returns to class having missed in which of two cupboards the teacher placed a toy plane. When asked where he will look for it, children using PAR claimed that he will choose the empty cupboard, because he didn't see the placement of the plane and therefore did not know where it was (and will "get it wrong" by looking in the incorrect cupboard).

acquisition (and theory change) is of course an empirical question, setting aside any a priori constraints (if indeed there are any). That is thus a question beyond the scope of this paper.

3.2 Knowing, Lying, and Getting it Right

According to Perner (1991, Chapter 7), "children's earliest use of 'know' emphasizes its link to successful action…" (p. 147). As he says, this is also a part of the adult use of 'know', as in 'knowing how to do something' or 'knowing where something is.'[108] This is the sense of 'know' involved in the PAR rules. However, in this stage of ToM development, it is a completely behavioral, non-representational term, and is devoid of other adult uses or aspects.[109] Thus the concept $KNOW_{PAR}$ can be separated from the adult conception of 'know,' even in this sense of leading to successful action. For example, in the adult concept, even if I don't know where my keys are, I still might accidentally choose the correct location in the first place I look. While for adults this would be a lucky guess of sorts, for children using PAR that means I know where they are. If knowing is simply "getting it right," or correct behavior, then there is no difference between an uniformed yet correct guess and a justified belief as long as both processes lead to

---

[108] Perner claims that this use of 'know' is the primary use for children until about four years of age. However, I cite evidence below that children don't acquire the representational concept of 'know' until later than that. Children in the RR stage of ToM development perhaps lack the concept of knowledge altogether (Bartsch & Wellman 1995 found that psychological uses of 'know'—versus mere conversational ones such as 'know what?' to introduce a topic—occur very rarely before age 3, and increase markedly during the fourth year). Even if they do have one, it must be different from $KNOW_{PAR}$.

[109] According to Perner (1991, Chapter 7)—drawing from Dretske (1981)—the other aspects of the adult concept KNOW are truth and "access to relevant information"; but these are unimportant for present purposes. Also, note that by "behavioral" I am not claiming that the child concept lacks any intentionality whatsoever; see Chapter 5 for more on this use of 'behavioral.'

successfully finding an object or accurately stating where an object is (or what properties the object has).[110]

Following this line of reasoning, a crucial distinction separating knowledge and lucky guesses is part of the adult representational and mental concept of knowledge but is lacking in $KNOW_{PAR}$. In (at that time ongoing) research discussed by Perner (1991, pp. 156-158), a piece of candy was placed in one of two differently colored boxes before subjects where asked, "Where is the candy?" In the *know* condition the candy was placed by the experimenter into one of the boxes while the subject watched. In the other two conditions an occluding screen was placed between the subject and the experimenter with the two boxes. The experimenter told the subject that he was placing the candy in one of the two boxes; however in the *guess-correct* condition candy was actually surreptitiously placed in both boxes and in the *guess-wrong* condition the candy was placed in neither box. Each subject saw all three conditions.

After being asked which box contained the candy and looking inside to witness either its presence or absence, the subject was asked, "Did you really know the candy was in there or did you just guess?" Fewer than 20% of the subjects between 4 and 5 years of age, and fewer than 40% of the subjects between 5 and 6.5 years of age, gave the correct response in all three conditions. Assuming that most of these children were in the PAR stage of ToM development, given their ages, this demonstrates that children using PAR (and the concept $KNOW_{PAR}$) do not

[110] Of course, if Williamson's view of knowledge (2000) is correct, then KNOW is simple (and not a composite containing BELIEVE, JUSTIFICATION, etc. as parts) and primary (BELIEVE and the like are variations of, and therefore parasitic upon, the concept KNOW). I tend to favor this view of knowledge myself, although I'm not prepared to argue for it here. Although it is not yet the full-fledged adult conception of knowledge, the fact that (according to the PAR hypothesis advocated here) children acquire the concept KNOW in the second stage of development, before they acquire BELIEVE in the third stage, may lend some support for Williamson's view (Williamson 2000, p. 33, footnote 7. See Chapter 7, Section 3.1 for discussion).

understand the difference between knowing something and merely guessing correctly. A child using PAR will say, in the *guess-correct* condition, that she really knew where the candy was because she successfully found the candy.

In fact, even children in the BR stage of ToM development may still take some time to fully grasp the difference between knowledge and guessing. Pillow and colleagues (Pillow 2002, Pillow et al 2000) found that it wasn't until 8 or 9 years of age that children were able to correctly judge another person's deductive inference as more certain than a guess. In a total of 5 studies, experimenters took two small toys (e.g. marbles) of different colors and placed them into opaque containers. The experimenter told the subject and a puppet that he was hiding e.g. "an orange marble and a blue marble." In the *looking* condition, the puppet looked into one of the tins and said, "This marble is blue." In the *inference* condition, the puppet looked into one of the tins and said, "The other marble is blue." In the *guessing* condition, the puppet pointed to one of the cans (without looking inside either) and said, "This marble is blue." Subjects were then asked to explain the puppet's belief (e.g. "Why does Bob [the puppet] think the other marble is blue?") and to rate the certainty of the puppet's belief by pointing to a scale which they had already been familiarized with ("How sure is Bob that the marble is blue? Can you show me with the arrow? Put it here [happy face at one end] if Bob is really, really sure, and put it here [sad face at the opposite end] if Bob doesn't know at all, and put it here [in the middle] if Bob is a little bit sure"). Most of the children (from preschool to 4th graders) did not consistently rate the puppet as less certain on guessing trials than on inference trials until about 8 or 9 years of age. For a child using PAR, the puppet only knows in the *looking* condition of these studies, because according to the first PAR rule if the puppet doesn't see the marble then she doesn't know what color it is.

Further evidence for this non-representational stage of ToM development comes from studies involving the concept of *lying.* Lying involves more than simply telling a falsehood. If I believe that my wife is at work, and thus tell my son that she's at work when he asks, "Where's Mom?", then I haven't lied. I haven't lied even if I make a mistake and she's actually on her lunch break at Subway. I have said something untrue, but it's not a lie because I didn't intend to deceive my son. I simply made a mistake. Lying involves knowing the truth, and intending to give someone a false belief about that fact. Of course, a false belief is a misrepresentation of reality that someone has in his mind. So if a child using PAR doesn't understand mental representation, then she can't fully understand lying. Lying for a child using PAR will be, just as ignorance is, a case of *getting it wrong*. There should be no difference between lying and mistakenly telling a falsehood in this stage of cognitive development, and indeed that is exactly what we find.

Piaget (1932/1965) interviewed a number of children ages 6-12 and told them stories involving children intentionally telling lies in order to get sweets, exaggerating due to fear, or making honest mistakes. He then asked subjects whether or not the children in the stories told lies or not, and how blameworthy ("naughty") their actions were. He found, in the first instance, that most young children don't distinguish between mistaken factual errors and lies. For example, one of the dialogues Piaget recites is with a 6-year-old boy named Clai (p. 143):

PIAGET: Do you know what a lie is?

CLAI: It's when you say what isn't true.

P: Is '2 + 2 = 5' a lie?

C: Yes, it's a lie.

P: Why?

C: Because it isn't right.

P: Did the boy who said that '2 + 2 = 5' know it wasn't right or did he make a mistake?

C: He made a mistake.

P: Then if he made a mistake, did he tell a lie or not?

C: Yes, he told a lie.

P: A naughty one?

C: Not very.

P: You see this gentleman [a student]?

C: Yes

P: How old do you think he is?

C: Thirty.

P: I would say 28.

[The student then says he is 36.]

P: Have we both told a lie?

C: Yes, both lies.

P: Naughty ones?

C: Not so very naughty.

P: Which is the naughtiest, yours or mine?

C: Yours is the naughtiest, because the difference is biggest.

P: Is it a lie, or did we just make a mistake?

C: We made a mistake.

P: Is it a lie all the same, or not?

C: Yes, it's a lie.

Cai (and other young children interviewed by Piaget) not only said that a mistaken falsehood is the same as a lie, but he also determined an untruth's moral blameworthiness using the dimension of realism. That is, the more incorrect a statement is, the more "naughty" it is. The children also considered the statement that has the worst results (i.e. the one that leads to unsuccessful action) the more blameworthy. In one study Piaget told subjects a pair of stories (making sure they could recite them from memory as a control question) before asking them which of the boys in the story was naughtier. In both stories (p. 149) a man stops to ask a boy directions to a particular street near the school where the interviews took place. In the first story, a boy who doesn't know the names of the streets very well tells him, "I think it is there," but the street is not there and the man seeking directions gets lost. In the second story, the man asks a boy who knows the names of the streets very well. However, the boy wants to play a trick on the man and shows him the wrong street. The man eventually finds his way again and does not end up getting lost.

Again, Piaget quotes dialogues with children ages 6, 7 and 8 who all claim that the first boy was the naughtiest because he made the man get lost. The important part of all of this for our purposes is that most of these young children don't distinguish between lying and making a mistake, which as we noted earlier is just what we would predict if the children are using PAR. Thus one child interviewed by Piaget, age 8, claims that if the man in the second story had also gotten lost, then both boys "would have been equally naughty" (p. 161), since they both lied. Lying is simply not telling the truth for subjects in the PAR stage of development.

Even with the current "Piaget Bashing" trend in developmental psychology which was noted at the beginning of this chapter (see Note 5), more recent studies confirm Piaget's original finding that young children deem all factually incorrect statements lies, even if the speaker is merely mistaken. Thus Wimmer et al (1984), who investigated children's understanding of lying in a series of 6 experiments, conclude, "The results of these experiments confirmed Piaget's claim that young children do not consider speaker intent in their use of the verb 'lying'" (p. 4).

In the first experiment, children watched 3 puppets—an older girl, a younger girl and a younger boy. In five different variations, the older girl answered the younger girl's question, who in turn reported to her brother (the younger boy). The test question was, "Did the small girl fib?"[111] Although the subjects performed excellently on control questions and in other situations, when the older girl sees a lion behind a wall but tells the young girl that it's a dog, and the young girl then tells her brother that the animal is a dog, the children answer that the small girl does fib. Of the 30 subjects between 4 ½ and 6 ½, only one answers correctly that she did not fib, even though they all answer correctly that she thinks the animal is a dog. (Interestingly, when the older girl tells the small girl that the animal is a lion, but she wants to trick her brother and tells him that it's a dog, but he mistakenly believes that the animal is a lion, subjects say that the small girl did not fib.)[112]

---

[111] The authors explain that they chose to use the expression 'fib' ('schwindeln' in German) rather than 'lie' ('luegen' in German) "since the situation was to be playful" (p. 5).

[112] In this instance, it's not entirely clear that adults wouldn't answer in the same way as the children. Perhaps 'to lie' is a success verb in the same way 'to con' is. However, I can certainly bring to mind counterexamples to this from everyday conversation. For instance, suppose my wife asks me whether I was working late or at the bar, when she knows I was at the bar. Even though it's unsuccessful when I tell her I was late at work, she might

Experiment 2 followed a similar paradigm except this time the experimenter enters with a box and tells the second character the contents of the box (either a piece of chocolate or a toy car), who then goes on to tell a doll. Sometimes the second character was played by a doll and sometimes the subject took the place of the second character. Neither the second character nor the doll she reports to are allowed to see inside the box until after the verbal report. The experimenter misinforms the second character, who in turn accurately reports to the doll what she was told by the experimenter. For example, the experimenter might tell the subject (playing the second character) that the closed box contains chocolate when it actually contains a toy car. The subject then reports to the doll that the box contains a piece of chocolate.

Once again, all but 2 of 20 subjects ages 4 ½ – 6 ½ answered that the second character was lying, and it didn't matter whether the second character was a doll or the subject herself. As the authors note, "One would have expected that in the Self condition subjects would have objected to even the slightest insinuation that they were lying (fibbing) when they themselves were led to a wrong belief by the experimenter. Instead, most of them even showed embarrassment when they saw that they had said something wrong in good faith" (p. 11).

Experiments 3 and 4 controlled for the possibility that the first character's lie influenced them to say that the second character was lying by using the false belief task paradigm (Wimmer & Perner 1983; see Chapter 2) instead. When Maxi returns after his mother has moved the chocolate from the red cupboard (where Maxi left it) to the green cupboard, Maxi

---

nonetheless complain that I'm lying. Thus I don't think *lie* is a factive or success verb such as *to know* or *to remember*.

tells his sister where the chocolate is. Sometimes Maxi wants to help his sister find the chocolate, and so he reports his false belief that the chocolate is in the red cupboard, which turns out to be incorrect. Sometimes Maxi wants to trick his sister, and so he tells her that the chocolate is in the green cupboard (where it actually is) instead of reporting his false belief that the chocolate is in the red cupboard. Once again, most of the subjects ages 4 ½ – 6 ½ said that Maxi was lying in the first condition, when he intends to tell the truth but makes a mistake, and said that Maxi was not lying in the second condition, when he intends to deceive his sister but accidentally tells her where the chocolate is actually located.[113]

To sum up this subsection, if *knowing* just means *getting it right* for the child in the PAR stage of ToM development (i.e. roughly ages 4 ½ to 6 ½), then there is no difference between what adults would call truly knowing and merely making a lucky guess. As research by Perner, Pillow, and their colleagues demonstrates, children do not make this distinction in the PAR stage of development.[114] Additionally, the full adult concept of lying (i.e. knowing the truth but attempting to deceive the listener by giving her a false belief) should be unavailable to a child using PAR, since the concept of lying requires a concept of mental representation which children in this stage of development lack. Therefore, for children at this stage there should be

---

[113] Experiments 5 and 6 are designed to disprove Piaget's (1932/1965) claim that young children base their moral judgment on the objective result of the action in question rather than on the subjective intentions behind that action, which he called Moral Realism. They used the same false belief task paradigm as Experiments 3 and 4, but also allowed subjects to reward (with 1, 2 or 3 stars) or punish (with 1, 2 or 3 black marks) Maxi's actions. Contra Piaget, 6-year-olds rewarded or punished actions based on intentions rather than the actual truth of their utterances. However, crucially for our purposes, even subjects who rewarded Maxi for his good intentions (trying to help his sister find the chocolate) nonetheless said he was lying (since his statement was not true).

[114] Children may not make this distinction until about age 8 or 9. This suggests that more than the concept of mental representation (which children acquire around 6 ½ years of age, as demonstrated by their ability to pass a two task battery of both the false and true belief tasks; see Chapter 2, Fabricius et al 2010, Hedger & Fabricius 2011) is required to fully grasp this difference.

no difference between what adults would call genuine lying and making an honest mistake which results in a falsehood. As research by Piaget, Wimmer, Gruber and Perner demonstrates, children call any falsehood a lie until around age 8, regardless of the speaker's intentions. This shows an important difference between the adult concept of knowledge and the concept of knowledge which is made use of in the PAR rules. These are ultimately two different concepts.

Of course, I grant that one could always resist this line of reasoning and argue that children in the PAR stage and adults share the same concepts KNOW, SEE, etc.; but that children simply understand less about them than adults do. However, I think it's a plausible position that adults and children in the PAR stage have distinct concepts. First, the types of errors that these children make are systematic. They are not due to any particular details of any one developmental psychology task, for instance, but occur across different contexts in a predictable manner. Second, the errors stem from a fundamental misunderstanding of the nature of mental states. Third, the sorts of mistakes these children make are quite surprising and stupefying from the adult point of view. Fourth, if the mistakes were due to some incomplete understanding, then education ought to help PAR children make fewer errors; but in actuality, giving them models to imitate, giving them feedback on earlier trials, etc. does not significantly improve their performance on these tasks. Fifth, the behavior of children and their performance on tasks undergoes perspicuous distinct stages of development, which suggests a fundamental shift rather than a completion of earlier incomplete understanding. For instance, children in the RR stage pass true belief tasks, children in the PAR stage fail them, and children in the BR stage pass them. Lastly, concepts are the constituents of thoughts (to have the concept CAT is to be able to think thoughts of which CAT is a part), and one way to explain the failure of children in the PAR stage is that they are unable to have thoughts which include representational mental states (where "representational" simply means roughly something in the mind which stands for things in the world). For these reasons, I find it plausible to think that children and adults have distinct mental state concepts. I also think that a theory-theory of mental state concepts is not as implausible as many philosophers have contended, which I argue in the following section.

3.3 The Child's Concept of Knowledge and Theories of Concepts

One way to classify among theories of concepts (which will be familiar to those who have read Fodor) is by separating the theories which count conceptual role as constitutive from those that do not. In other words, are concepts partly individuated by the role they play in larger cognition? Should things such as inferential role and abilities to sort things which fall under the concept from those that do not be considered part of the content of concepts? Theories such as Fodor's (1998, 2004) *Concept Atomism*, where what is constitutive of the basic concepts is completely divorced from their role in cognition or their relations to other concepts, or Edwards's (2009, 2010, 2011) *Concept Referentialism*, where the content of a concept is exhausted by its referent, will answer these questions in the negative. A *theory-theory* of concepts, such as the one advocated here for the concept KNOW during the PAR stage of cognitive development, says that a concept is partly individuated by the role it plays in a larger theory, and hence this will be one of the many theories of concepts which are on the opposite side of the fence from Fodor.

Of course, as Edwards (2010) points out, even if conceptual role is not part of the content of a concept, unless something goes terribly wrong conceptual role should nonetheless be "in sync" (as he puts it) with its content. Even if the content of CAT is simply that it refers to cats, normally possession of the ability to sort cats from non-cats will be a reliable indicator of possession of the concept CAT; those who think about cats will infer that they are animals and typically chase mice; and so on.[115] The critical test for deciding which theory of concepts is correct, then, is whether content can be in principle divorced from conceptual role. Fodor (1998, 2004, 2008) argues that it is certainly possible that one could possess e.g. the concept CAT and yet fail to recognize cats, fail to be able to sort cats from non-cats, fail to infer anything about cats, and so on. In order to possess the concept CAT, the only ability required is the ability to think about cats. For the case of CAT, I'm inclined to agree with Fodor, that even though the hypothetical situation he presents is a practical impossibility, the content of the concept CAT is logically distinct from the conceptual role of CAT.

However, I find the case of KNOW presented here (and indeed, *all* mental state concepts acquired during cognitive development) to be importantly different from the case of CAT. In particular, the PAR child's concept KNOW is not separable from the inferential rules which she uses to recognize cases of knowledge. These rules, and thus conceptual role, are in fact what constitutes the concept of KNOW for 4- and 5-year-olds, as shown from our definition above. The fact that $KNOW_{PAR}$ does not involve mental representation at all, the fact that I take

[115] Edwards's (2010) example of "talking on the phone," and the language of heterogeneous causal processes "implementing," "realizing," or "instantiating" the reference relation which he takes to be constitutive of concepts (p. 101), readily brings to mind an analogy with *functionalism* in the philosophy of mind. His real move is to place the individuation of concepts at a higher level of abstraction in a certain sense.

mental representation to be a crucial component of the adult concept of KNOW, and the fact that the two concepts are thus made incommensurable, make KNOW an example of the sort of conceptual change envisaged by Susan Carey (1985, 2010).[116] The fact that, according to the present view, the conceptual role of KNOW_{PAR} is constitutive of the concept, opens up the view to Fodor's criticisms. In the next section, we shall see whether or not we can answer one of those criticisms, viz. the challenge of *shareability*.

4. The Shareability Objection

Shareability is the term I shall use (following Edwards, 2010; see note 5 above) for one of Fodor's three constraints on a theory of concepts.[117] The basic idea (see e.g. Edwards, 2010, p. 95; Fodor, 1998, pp. 28-34) is that in order for intentional psychological generalizations to be true, it is necessary that different people think the same thoughts on occasion. A fortiori, people must be capable in principle of thinking the same thoughts, and hence the constituents of those thoughts, i.e. concepts, must be capable of being tokened by different people on different occasions. In a word, concepts must be *shared*. Furthermore, surely anyone who takes the science of psychology at all seriously is committed to there being true generalizations about our intentional states. Even more basically, it would seem intuitively obvious that in order to communicate with each other or understand what others have written, etc. we must share a

[116] Is 'incommensurable' too strong? Use of this term in the concept literature of course refers to Kuhn's (1962) usage, and its use in Philosophy of Science is similarly contentious. I'd probably be willing to trade 'incommensurable' for a weaker term, but it's certainly true that children and adults do a lot of talking past each other, and it takes a great deal of understanding of ToM development to even begin to understand what children at this stage mean by "know." I'm not sure I completely grasp it even now, keeping my definition in mind and maintaining extensive effort. Like Descartes at the end of the First Meditation, I find myself slipping back to my old opinions. For these reasons as well as the common usage of the term in this context, I'm willing to risk hyperbole in order to emphasize the vast differences between the child's mental state concepts and our own.

[117] At least, 3 by Edwards's (2010) count (p. 95); there are 5 constraints by Fodor's (1998) count (p. 34).

large number of concepts. When I visit San Francisco for the first time and ask about a good restaurant, it must be assumed that whomever I am asking means the same thing by "restaurant," "food," etc. that I do, and that requires that we are both tokening concepts of the same type FOOD, RESTAURANT, and so on.[118]

Of course, PAR theorists and adults do *not* share the concept KNOW, because this word maps onto different concepts for the child than it maps onto for the adult. At this point, a concern naturally arises: How are adults and children able to communicate? In particular, how are they able to communicate about knowledge? Carey (1992)'s answer to this is, I think, spot on. She says (p. 99; cf. Carey, 2010, p. 377 ff.):

> Incommensurability does not require complete lack of communication. After all, the early oxygen theorists argued with phlogiston theorists, who were often their colleagues or teachers. Locally incommensurable languages can share many terms that have the same meaning in both languages. This common ground can be used to fix referents for particular uses of non-shared terms, e.g. a use of "dephlogistonated air" to refer to oxygen enriched air. With much common knowledge the two sides can have genuine disagreements about the nature of dephligistonated air. Anyway, it is an empirical question just how well adults understand preschool children [and vice versa].

[118] Some argue that this shows not that the concepts we token must be of the exact same type, but only that the content of those types must overlap sufficiently to facilitate communication and intentional generalizations. Setting aside the problem of giving some analysis of whatever *sufficient similarity* amounts to, there remains Fodor (1998)'s extended argument that these solutions to the problem won't work (pp. 30-34). I will ignore this idea for the remainder of this chapter, mostly because my proposed solution need not make any such appeal.
This particular example may bring about worries for those who think propositions involving matters of taste are relative, since it would seem to follow from this view that different speakers might mean different things by '*good* food.' Personally, I am not at all sympathetic to this (or any other) kind of relativism, although I'm not prepared to argue my point here. If you do have this concern, feel free to substitute the example of asking for a restaurant that serves seafood.

So how can adults teach children about the concept KNOW? Perhaps the fact that the adult and the child share enough other concepts, combined with the fact that they share the word 'know,' allows children to use that word as a kind of placeholder and "bootstrap" the adult concept (Carey, 2010). However, as Carey urges above, how well adults and children communicate about knowledge is an empirical question. Numerous theory of mind experiments in the developmental literature suggest that there may be a lot more talking past each other than adults typically assume. What's more, my hypothesis is that adults don't teach children a darned thing about the concept KNOW. More on this below. For now, I take the initial worry to be answered.

Another problem arises at this point, however. Namely, how is it that one person's mental state concept is the same as another's, even at the same stage of cognitive development? After all, different people know and understand different things. People must share concepts, but it seems obvious that people exhibit epistemological variations. After all (the criticism goes), people can believe almost anything, no matter how crazy. If mental state concepts are partly individuated by the theories in which they take part, wouldn't it be likely that each person's theory has some slight differences with each other person's? If so, then it would seem to follow from the theory-theory view that each person's concept KNOW is different. The familiar problems of holism begin to emerge at this point (Fodor & Lepore 1992, 2002).[119]

---

[119] I recently saw a talk by John Hawthorne (at the retirement event for José Benardete, Syracuse University, 13 December 2011) in which he (implicitly) argued that these problems are not escaped by meaning atomists. If physicalism is true, then semantic meaning supervenes on some physical base. That physical base can change across dimensions such as time and space. Semantic meaning must therefore change over time and space.

To begin our search for a solution to this problem, we need to reconsider the history of science analogy that psychologists have used as a motivation for the theory-theory of concepts (Carey, 1985, 1992, 2010; Gopnik, 1988). Although supporters of this view have undoubtedly done important work, in general I am not a fan of this model, for two main reasons: First, an analogy is only helpful when it allows us insight into a phenomenon we don't understand (such as the mind) by comparing it to something that we have a better grasp of (such as the digital computer). In this particular case, we don't yet have a good enough grasp of philosophy of science for it to shed any light on concept acquisition. There is no consensus on an analysis of how theories work in science, or how transitions between dominant theories take place. It seems to me that we are simply trading one mystery for another when we make an analogy between the acquisition of concepts in childhood and the history of science.[120]

Second, and more to the point for present purposes, this analogy may have caused developmental researchers to ignore some important *differences* between the progression of science and concept acquisition. For instance, in science there are typically leaps made by some genius such as Newton, Einstein, Darwin or Turing. They come up with some new, innovative, and important ideas, and then have the job of explaining this to the rest of the science

(Memory fails me, but I think that there's a Dante quote about the language subtly changing in the villages across Italy, so that each person can understand someone in the nearest village, but not a person from a village further along the trail.) Thus, even if atomism can answer how someone else in the same room can share concepts, it leaves unanswered the problem of how those same concepts remain constant over time. This seems to be a serious concern, but one which we don't have time to consider further here. My hunch is that Fodor had more credence in concept atomism (not to mention irreducibility of the sciences) than he did in physicalism (this much is suggested by remarks in Fodor & Lepore 2002, p. 10 e.g.). Sadly, we can no longer ask him if my hunch is correct.

[120] Alison Gopnik (1988) is sensitive to this, but for some reason which I don't fully grasp finds it a virtue of the theory rather than a vice (p. 209). Fodor (1998) clearly agrees with me here; he notes (with his characteristic acerbic wit), that "if you find the idea that a scientific theory-change is a paradigm shift less than fully perspicuous, you will also be uncertain what exactly it is that the ontogenetic analogy asserts about stages of conceptual development" (p. 113).

community. The general population, if they're fortunate, also learn about the theory in time. (Maybe I'm less intelligent than the average Joe, but I still have difficulty fully grasping Einstein's relativity theory, even after reading a number of explanations aimed at laymen.)

On the other hand, folk psychological theories—including the ones held at early stages of cognitive development—are not like this at all. They are not the product of an original insight of some genius, but are instead found everywhere. A theory of mind seems to be a universal human phenomenon (with the possible exception of some people with autism spectrum disorder—see e.g. Baron-Cohen 1995). Furthermore, they don't seem to be the kind of things that are learned. RR and PAR theories are thought to be wildly false by adults, so there is no reason that they would be teaching these theories to their children. Also, a meta-analysis of 178 theory of mind studies (Wellman, Cross, and Watson 2001) found that neither education nor country of origin can improve 3-year-olds' performance on the false belief task to higher than chance. Thus, culture and upbringing are not relevant factors in a child's performance on theory of mind tasks.

Also, it has been widely noted that attempts to teach children at an earlier stage of cognitive development has absolutely no effect on their ability to pass ToM tasks. Flavell et al (1995) found that extensive pre-training periods concerning what it means to think about something did not improve the performance of 5-year-old subjects on tasks where they were asked to determine which of a choice of objects they were thinking about a minute ago. For instance, in one study (Study 11) the five-year-olds were seated at a table and asked which of two crayons on the table is the longer one. They then immediately moved to a chair on the opposite side of the table, and were asked which objects they were thinking about while seated in the first chair—the two crayons, a plastic doll which was hidden from view in an envelope while the child was seated in the first seat, or the experimenter's eyeglasses. Although an adult experimenter participated along with the subjects and modeled correct answers to the questions, only 19% of the 5-year-olds correctly chose the crayons and not the irrelevant objects (one of which was completely hidden from view until the questioning). A pre-training period, involving using a flashlight as a metaphor, did not significantly improve overall performance (Study 12), and in fact in some cases the performance was actually worse. Johnson & Wellman (1982) asked children ages 3 to 14 whether the brain was used in various mental activities, including remembering, seeing, knowing the alphabet, and many others. A unit on the brain included in the 5th grade science curriculum had absolutely no effect on the results (as reported in Carey, 1985, pp. 48-51).[121] Further examples abound in the literature.

This sort of evidence leads to the hypothesis which I wish to propose here, which would also serve as a solution to the shareability problem. The answer to the problem is that concepts are shared because these theories are *innate*. Innateness may be difficult to specify, but for our purposes we may perhaps say that a cognitive mechanism is innate just in case it is the result of some biological phenotype. The key point is to exclude those mechanisms which are learned and those which are acquired by some sort of accident (such as getting hit on the head by a brick).[122] Innate cognitive mechanisms will ultimately have an *evolutionary* explanation for their

---

[121] This phenomena is not limited to the understanding of mental concepts; Carey (2010, pp. 436-445) reports that extensive math and science education, from elementary to high school, has not helped a vast number of people understand math and science concepts such as decimals, fractions, weight, and mass.

[122] See Fodor (2008) pp. 131-132 for an amusing list of possible processes which might be called concept acquisition. A successful analysis of 'innate' would rule out all of these processes except for the one which Fodor

existence, rather than any learning process or incidental circumstances. The argument for nativism concerning mental state theories is that (1) these theories (RR, PAR, and BR) are universally found among all human beings regardless of their culture, upbringing or education, (2) these theories do not appear to be the result of any learning process, and (3) the sole major factor in the acquisition of these theories seems to be the *age* of the subject (e.g. Wellman et al, 2001). Admittedly, we currently lack any sort of model in cognitive science for understanding what sort of mental mechanism an innate theory would be like,[123] but the empirical evidence points towards nativism as the best explanation of the nature and existence of these theories, and that also happens to provide a solution to the shareability worry.

Interestingly, this hypothesis was initially put forward by no less than Jerry Fodor himself, in an overlooked but ingenious argument in the "Creation Myth" epilogue to *Psychosemantics* way back in 1987. There he argues that what makes the higher intelligence of humans possible is the ability to accumulate and transmit information through language "(and other cultural artifacts)" (p. 130). "Ideally," he says, "each *individual* should be the beneficiary

---

(1975) initially endorsed for all concepts. Importantly, however, I wish to argue only for nativism concerning mental state concepts such as KNOW.
Of course, carefully distinguishing innate from acquired characteristics in psychology is notoriously difficult (see Griffiths, 2009, for discussion). Chomsky's account of language has been highly influential to contemporary discussions of nativism in psychology and the philosophy of psychology. His "poverty of the stimulus argument" is roughly that the small amount of exposure and feedback available to young children concerning grammaticality of language strings is insufficient to explain the vast knowledge of grammar exhibited by their linguistic behavior, including frequent production of novel grammatical strings (Chomsky, 1959). Here the relevant contrast seems to be between knowledge which is innate and knowledge which is the result of a cognitive learning process (cf. Fodor, 1975). This is admittedly a difficult issue which I can't help much with here, but I hope my simple definition (along with pretheoretic intuitions) suffice to motivate the view I propose. (Personally, I am waiting for a genius to come along and explain it to the rest of us. In the meantime please excuse the vagueness of my proposed definition.)

[123] Core cognition (Carey 2010) and modularity (Fodor 1983) seem to come close, but will have to be modified. (Ultimately I think Chomsky's 1965 theory of universal grammar comes closest; more on that below.) Also, although this is extremely speculative, if some theories are indeed innate, perhaps that could that be the start of an answer to Fodor's (2000) dilemma about where abduction comes from.

of the learned adaptations of *all* of its conspecifics, cohorts, and predecessors alike" (p. 129). Since it's not possible for this transmission of information to occur through heritable traits, then it's done through language and other cultural processes—the next best thing. However, this process presupposes a culture, and a culture in turn "presupposes a social animal, one that is capable of integrating its own behavior with the behavior of others of its kind… You do not want to spend your life reinventing the wheel. But neither do you want to have to spend it learning enough about your neighbor's psychology to permit you to exploit his expertise" (p. 131). The ability to be social with creatures as cognitively complex as human beings requires a folk psychological *theory of mind*. I can't meet you at your office without making some assumptions about your beliefs and desires and how they lead to behaviors such as going to such-and-such location at such-and-such time.[124]

A further constraint on what we can learn is the relative brevity of the human lifespan. Fodor explains that this process of coming to understand "must be achieved *rapidly* compared to the length of an individual life. There is … no use designing a social organism with a long prematurity if most of its apprentice years have to be spent learning the commonsense psychology of its species" (p. 132). And another problem is that the intelligence of human beings makes the behaviors of which they are capable very complex and complicated, so that a

[124] This is similar to the *Cultural Intelligence Hypothesis* of Tomasello and colleagues (Herrmann et al 2007, Tomasello 2009), that humans' evolved social-cooperative and ToM skills, which enable transmission of knowledge through culture, are the basis of cognitive skills which appear far superior to those of any other species. This hypothesis also has some empirical support. To take one example, an extensive study (Herrmann et al 2007) found that 105 2 ½ -year-olds did not perform significantly better than 106 chimpanzees on tasks testing understanding of space, quantity and causality (both groups performed slightly better than 32 orangutans in a total of 78 trials), although they performed about twice as well as chimps and orangutans on tasks testing social learning, communication and theory of mind.

simple mechanical solution (Fodor gives the example of Stickleback mating procedures, pp. 131-132) will not work. What is needed is a fairly robust *theory*. Fodor concludes (bold added for emphasis):

> Here is what I would have done if I had been faced with this problem in designing *Homo sapiens*. **I would have made knowledge of commonsense *Homo sapiens* psychology *innate*; that way nobody would have to spend time learning it.** (p. 132)

That is Fodor (1987)'s a priori argument that a folk psychological theory of mind is innate.

Fodor (1987) also gives an empirical argument to the best explanation for the nativism of theory of mind not unlike my own, above. He urges that, "The empirical evidence that God did it the way I would have isn't, in fact, unimpressive" (p. 132), and he provides three reasons in support of his abduction. First is the same universality I noted above (see the epigram to this article for the quote). Second is the lack of a rival hypothesis (p. 133). This is perhaps akin to my point that these theories don't appear to be learned. The third point must be amended somewhat. Fodor hypothesized that we would find demonstrations of children's understanding of the mind at earlier and earlier ages once experimental techniques improved. This is certainly understandable given the progression of developmental psychology in 1987. As explained earlier, that was typically the flavor of the "Piaget bashing" experiments which were taking place during the 1970's and 1980's. However, time has not come to prove Fodor's point in the case of theory of mind development. In fact, the state of the art—the PAR hypothesis of Fabricius and colleagues—has shown that theory of mind development is *more* protracted than previously thought, not less so.[125] Even so, given the age factor, the succession of these cognitive stages may simply be a result of cognitive mechanisms (and theories) "coming online." That is, they may simply be due to maturity of the neurological system coupled with some necessary social interaction.

Lastly, while I'm in a speculative mood, here's another radical hypothesis which has always seemed plausible to me: KNOW is an innate, primitive concept that comes online along with the BR theory. That's why adults have such strong intuitions about when someone knows, but 2500 years of philosophy hasn't been able to come up with a good analysis of what knowledge is.[126] There is a striking analogy here with intuitions concerning the grammaticality of strings of language and Chomsky's theory of universal grammar (Chomsky 1965). Native speakers are very good at determining which strings are grammatical and which are not, but very bad at explaining what makes a given string ungrammatical. Similarly, adults are largely in agreement about what cases constitute knowledge and which do not, but are very poor at explaining why. Furthermore, these rules of grammar which Chomsky and other linguists argue are innate, are similar in complexity to what a theory of knowledge might be like. So perhaps there is a precedent for this sort of hypothesis in cognitive science after all.

---

[125] Some infant researchers currently claim otherwise, but see Hedger and Fabricius (2011) for a critical review.

[126] The once much discussed finding that Asians don't share Gettier intuitions about knowledge (Weinberg et al 2001) would, of course, throw a monkey wrench into this hypothesis, *if it were genuine.* Luckily for me, this initial study doesn't appear replicable (Nagel 2012, Seyedsayamdost 2015).

References

Astington, J. W. (1993). *The child's discovery of the mind.* Cambridge, MA: Harvard University Press.

Astington, J. W. & Gopnik, A. (1988). Knowing you've changed your mind: Children's understanding of representational change. In Astington et al (1988), pp. 193-206.

Astington, J. W., Harris, P. L. & Olson, D. R., Eds. (1988). *Developing theories of mind.* New York: Cambridge University Press.

Bartsch, K. & Wellman, H. M. (1995). *Children talk about the mind.* Oxford, England: Oxford University Press.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Carey, S. (1988). Conceptual differences between children and adults. *Mind and Language, 3,* 167-181.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey and R. Gelman, Eds., *The epigenesis of mind: Essays on biology and cognition.* Hillsdale, NJ: Erlbaum, 257-291.

Carey, S. (1992). The origin and evolution of everyday concepts. *Minnesota Studies in the Philosophy of Science, 15,* 89-128.

Carey, S. (2010). *The origin of concepts.* Oxford, England: Oxford University Press.

Chen, Y., Su, Y., & Wang, Y. (2015). Young children use the "ignorance = getting it wrong" rule when predicting behavior. *Cognitive Development, 35,* 79-91.

Chomsky, N. (1959). A review of Skinner's *Verbal behavior. Language, 35*, 26-58.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: The MIT Press.

Dennett, D. (1978). A cure for the common code? In *Brainstorms: Philosophical essays on mind and psychology.* Cambridge, MA: Bradford Books.

Doherty, M. J. (2009). *Theory of mind: How children understand others' thoughts and feelings.* Hove, England: Psychology Press.

Dretske, F. I. (1981). *Knowledge and the flow of information.* Cambridge, MA: MIT Press.

Dutton, D. & Aron, A. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology, 30,* 510-517.

Edwards, K. (2009). What concepts do. *Synthese, 170,* 289-310.

Edwards, K. (2010). Concept referentialism and the role of empty concepts. *Mind & Language, 25,* 89-118.

Edwards, K. (2011). Higher-level concepts and their heterogeneous implementation: A polemical review of Edouard Machery's *Doing without concepts*. *Philosophical Psychology, 24,* 119-133.

Fabricius, W. V., Boyer, T., Weimer, A. A., & Carroll, K. (2010). True or false: Do five-year-olds understand belief? *Developmental Psychology*, *46*, 1402-1416.

Fabricius, W.V., & Imbens-Bailey, A.L. (2000). False beliefs about false beliefs. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning about the mind* (pp. 267-280). Hove, England: Psychology Press.

Fabricius, W. V., & Khalil, S. L. (2003). False beliefs or false positives? Limits on children's understanding of mental representation. *Journal of Cognition and Development, 4*, 239-262.

Flavell, J. (1974). The development of inferences about others. In T. Mischel, Ed., *Understanding other persons.* Oxford, England: Blackwell, 66-116.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1983). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development, 51*, pp. 1-89.

Flavell, J. H., Green, F. L. & Flavell, E. R. (1995). Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development, 60*, pp. 1-96.

Fodor, J. A. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

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

Fodor, J. A. (1985). Fodor's guide to mental representation. *Mind, 94,* 76-100.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind.* Cambridge, MA: The MIT Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong.* Oxford, England: Clarendon Press.

Fodor, J. A. (2004). Having concepts: A brief refutation of the 20th century. *Mind & Language, 19,* 29-47.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology.* Cambridge, MA: The MIT Press.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited.* Oxford, England: Oxford University Press.

Fodor, J. A., Garrett, M., Walker, E., & Parkes, C. (1980). Against definitions. *Cognition, 8,* 263-367.

Fodor, J. A., & Lepore, E. (1992). *Holism: A shopper's guide.* Oxford, England: Blackwell.

Fodor, J. A., & Lepore, E. (2002). *The compositionality papers*. Oxford, England: Oxford University Press.

Friedman, O., Griffen, R., Brownell, H., & Winner, E. (2003). Problems with the seeing = knowing rule. *Developmental Science, 6*, 505–513.

Gallois, A. (1996). *The world without, the mind within: An essay on first-person authority.* Cambridge, England: Cambridge University Press.

Gazzaniga, M. S. (1985). *The social brain: Discovering the networks of the mind.* New York, NY: Basic Books.

Gazzaniga, M. S. & LeDoux, J. E. (1978). *The integrated mind.* New York, NY: Plenum Press.

Gelman, R. & Baillargeon, R. (1983). A review of some Piagetian concepts. In P. H. Mussen, Ed., *Handbook of child psychology, 4th edition:* J. H. Flavell & E. M. Markman, Eds., *Volume III: Cognitive development.* New York, NY: Wiley & Sons, 167-230.

Gertler, B. (2011). *Self-knowledge.* New York, NY: Routledge.

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

Gopnik, A. (1988). Conceptual and semantic development as theory change. *Mind & Language, 3,* 197-217.

Gopnik, A. & Astington, J. W. (1988). Children's understanding of representational change and its relation to their understanding of false belief and the appearance-reality distinction, *Child Development, 58,* 26-37.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7,* 145-171.

Gopnik, A., & Wellman, H. M. (1994). The "theory theory." In L. Hirschfield and S. Gelman, Eds., *Domain specificity in culture and cognition*. New York: Cambridge University Press, 257-293.

Griffiths, P. (2009). The distinction between innate and acquired characteristics. In E. Zalta, ed., *Stanford Encyclopedia of Philosophy, Fall 2009 Edition.*

Hedger, J. A. (2016). Perceptual access reasoning: Developmental stage or System 1 heuristic? *Phenomenology and the Cognitive Sciences, 15,* 207-226.

Hedger, J. A., & Fabricius, W. V. (2011). True belief belies false belief: Recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development. *The Review of Philosophy and Psychology, 2,* pp. 429-447.

Herrmann, H., Call, J., Hernandez-Lloreda, M. V., Hare, B. & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317,* 1360-1366.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition, 109*, 224 – 234.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis and S. Laurence (Eds.), *Concepts: Core readings*, pp. 3-81. Cambridge, MA: MIT Press.

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual precepts of others. *Child Development, 45*, 357–366.

Nagel, J. (2012). Intuitions and experiments: A defense of the case method in epistemology. *Philosophy and phenomenological research, 85,* 495-527.

Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: The MIT Press.

Piaget, J. (1932/1965). *The moral judgment of the child.* Trans. Gabain, M. New York: Free Press. [Original published in 1932.]

Piaget, J., & Inhelder, B. (1974). *The child's construction of quantities.* Trans. A. J. Pomerans. London, England: Routledge and Kegan Paul.

Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology, 47,* 116-129.

Pillow, B. H. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child Development, 73,* 779-792.

Pillow, B. H., Hill, V., Boyce, A., & Stein, C. (2000). Understanding inference as a source of knowledge: Children's ability to evaluate the certainty of deduction, perception, and guessing. *Developmental Psychology, 36,* 169-179.

Schachter, S. & Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review, 69,* 379-399.

Seyedsayamdost, H. (2015). On normativity and epistemic intuitions: Failure to replicate. *Episteme, 12,* 95-116.

Shoemaker, S. (1988). On knowing one's own mind. *Philosophical Perspectives, 2,* 183-209.

Tomasello, M. (2009). *Why we cooperate.* Cambridge, MA: MIT Press.

Weinberg, J. M., Nichols, S. & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics, 29,* 429-460.

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.

Williamson, T. (2000). *Knowledge and its limits.* Oxford, UK: Oxford University Press.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: Belknap Press.

Wimmer, H., Gruber, S., & Perner, J. (1984). Young children's conception of lying: Lexical realism—moral subjectivism. *Journal of Experimental Child Psychology, 37,* 1-30.

Wimmer, H. & Hartl, M. (1991). Against the Cartesian view on mind: Young children's difficulty with own false beliefs. *British Journal of Developmental Psychology, 9*, 125-138.

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.

CHAPTER SEVEN

Further Directions:

The next phase of perceptual access reasoning research

1. Introduction

This monograph essentially represents the first decade of an ongoing research program in the Philosophy of Psychology. This chapter includes an outline of projects for further development of the Perceptual Access Reasoning (PAR) theory (Section 2) and useful ways to apply the theory to solve more problems in philosophy of mind (Section 3). In Section 4 I provide a more detailed outline for a future project exploring the phenomenal experience of young children who are using PAR, entitled *What is it like to be a 5-year-old?*

2. PAR Theory Development

As I hope to have demonstrated in this monograph, the PAR hypothesis has been developed into an extensive theory which makes better explanations of the data on belief and perception than the other, currently more more popular, theories. It also makes some interesting predictions, many of which have been demonstrated empirically, but many more of which still need to be tested. Brief outlines of more specific projects in this area follow.

2. 1 Desire and other mental states from PAR perspective

In Chapter 2 I gave an account of belief and how this concept develops in children, based upon the framework of the PAR hypothesis. In Chapter 3 I demonstrated that the data gathered on children's understanding of perception and its development, influenced by the Flavell lab's four levels model and appearance/reality distinction, is consistent with the PAR hypothesis. Belief and perception have always been of central importance to philosophers, and of course are prominently featured in the two PAR rules (see Chapter 2). One issue that needs further work is exactly what the "knows" in the PAR rules really consists of. It cannot be full-fledged belief, since this concept is not acquired until the BR stage. (See Chapter 5 and Section 3.1 below for more on this topic.) The no belief task discussed in Chapter 2 could use replication, and there may be other tasks involving belief that could strengthen the case for the PAR hypothesis as well. It would also be a good future project to design one or more perception tasks in order to provide some clarity on children's understanding of perception and how it develops as well (see also Section 2.6, below).

Of course, mental phenomena involve much more than believing and perceiving. Another key mental state for Theory of Mind (ToM) development is *desire*. ToM as discussed by psychologists is basically what philosophers have called *folk psychology* (see e.g. Churchland 1981, Davidson 1963, Dennett 1981, Fodor 1987, Horgan & Woodward 1985, Lewis 1972, Ramsey, Stich & Garon 1990). Philosophers have noted that a person's beliefs and desires are central, perhaps sufficient, to explain a person's behavior, using something like Aristotle's practical syllogism (*Nichomacean Ethics*): Joe wants a beer (desire), thinks that there is some in the refrigerator (belief), and this explains why he is opening said refrigerator (action). Thus, a better understanding of children's understanding of desire and its development from the point of view of the PAR theory is needed. There is already some evidence for an early non-representational concept of desire, something like a simple, behavioral environmental drive towards an object (Wellman & Woolley 1990). It seems likely that there will be a shift to a concept of representational desire around age 6 when children hit the Belief Reasoning (BR) stage, to accompany the acquisition of representational concepts of belief (Chapter 2) and Perception (Chapter 3).

Desire presents a unique challenge when designing tasks to test between the use of a representational concept of desire and non-representational concept of a drive towards an object. Both belief and perception can fail to accurately represent the world, and these non-veridical possibilities are exploited in the tasks testing them. For instance, in the (standard) *false belief task* the question is whether children understand that Maxi has a non-veridical representation of the location of the chocolate (see Chapter 2), and in the *appearance/reality task* the question is whether children understand that someone can have a non-veridical representation of a sponge (see Chapter 3). The key to these tasks and their variations is that without a concept of representational mental states, non-veridical mental states do not make sense.

However, when it comes to desire, people can desire things that the world does not currently contain. This has to do with what Anscombe (1957) called *direction of fit*—whereas beliefs and perceptions need to be adjusted so that they align correctly with the world in order to be accurate, desires are satisfied when the world is altered to fit the representation. My desire for a Mercedes is satisfied when I am able to change the world so that my name is on the title of one. I suppose a non-veridical desire would be a desire that the subject does not genuinely have, and so desire is much more difficult to satisfactorily demonstrate experimentally.[127] Whereas beliefs are descriptive representations of how the world is, desires are prescriptive representations of how the world ought to be.

Researchers have often thought of ToM task demands in terms of the ability to hold two conflicting mental states in the mind at once rather than in terms of mental representation, and so Repacholi and Gopnik (1997) tested whether children would offer an adult broccoli instead of crackers when that adult displayed a preference for them, even though the subject herself preferred the crackers. Although success on this task indeed demonstrates some understanding of desire, it does not seem to distinguish representational desire from non-representational desire. This is analogous to Level 1 understanding of perception in Flavell's model discussed in Chapter 3. Passing this task only involves understanding that people can have desires that are different from one's own (i.e. moving past *egocentrism* in Piagetian terms), but each of these desires might simply be behavioral drives for objects in the immediate environment. A new type of task is required to test for a PAR stage of desire.

Perhaps one way would be to test whether a child subject understands that a desire persists through situation change similarly to beliefs; for example that a person who desires X will continue to desire X even when it is not present in the immediate environment (i.e. even when perceptual access is lacking), and will attempt to alter the world to make X present if it is within the person's ability to do so. For instance, tell the child that Maxi desires chocolate when

[127] Of course, this gloss of "a desire that the subject does not genuinely have" would need a lot more explication, and can't be exactly right. There is a lot of literature in psychology e.g. concerning subconscious desires that subjects are not aware of, and so on. The main point I am trying to make here is that the veridical/ non-veridical distinction which belief tasks use to test subjects doesn't map neatly onto desires. However, I don't want to be too quick and say there is no such distinction for desire, since people have come up with unexpected similar distinctions for other sorts of non-descriptive content, such as David Kaplan has done for expressives.

both crackers and chocolate are present, and have Mom put the chocolate away while Maxi is outside playing. When Maxi returns, the chocolate is now occluded but he retains perceptual access to the crackers. Have Maxi walk toward the crackers, and the test question might be, "What does Maxi want? Does he want the chocolate? Or does he want the crackers?" At any rate, further work needs to be done in order to adequately test for a PAR stage of desire.

Lastly, although desire would be the logical next mental state to study from the PAR perspective given its centrality in folk psychological prediction and interpretation, there are of course many other types of mental states—imagining, considering, wondering, understanding, daydreaming, boredom, confusion, irritation, obsession etc.—and some of these may be interesting to explore through the lens of the PAR hypothesis.

2.2 Testing adults' eye gaze on true and false belief tasks

The first major challenge to what I call the *Traditional View* of ToM development in children, viz. that children progress directly from RR to BR at about the age of four (see e.g. Gopnik & Wellman, 1992; Perner, 1988, 1991; Saxe, Carey, & Kanwisher, 2004; Wellman et al, 2001; see Chapter 2 above for discussion), was the finding that 3-year-olds appear to implicitly pass the standard, 2-location false belief task by showing unconscious anticipatory looking to the correct location (Clements & Perner, 1994; Garnham & Perner, 2001; Garnham & Ruffman, 2001; Ruffman, Garnham, Import, & Connelly, 2001).

However, as explained in Chapter 2 above, the standard 2-location false belief task is unable to differentiate between the use of BR or PAR. In other words, subjects may of course pass by reasoning that Maxi retains a mental state of belief representing the location of where he left the chocolate—which persists while he is outside regardless of what happens in the room during his absence and regardless of the true location of the chocolate. However, they may also pass by using the PAR rules, and reasoning that when Maxi returns he is in a new situation, and in this new situation Maxi does not have perceptual access to the chocolate. Hence, Maxi doesn't see the chocolate, therefore doesn't know where the chocolate is, and will get it wrong by choosing the location where the chocolate is not currently located.

The eye gaze studies to date therefore inherit this confound from the classic false belief tasks on which they are based (Clements & Perner, 1994; Garnham & Perner, 2001; Garnham & Ruffman, 2001; Ruffman, Garnham, Import, & Connelly, 2001), and thus they cannot accurately determine whether subjects are using BR or PAR in order to look at the correct location. These studies also suggest that there are simultaneous psychological mechanisms active in 3-year-olds, along the lines of a Dual Systems model such as Kahneman (2011). There appears to be an implicit modular mechanism which would explain anticipatory looking to the correct location, as well as the explicit use of RR which explains their verbal response of the incorrect location.

Hedger & Fabricius (2011) conjecture that *Rule A* is this modular, subconscious mechanism. Rule A is a condensed version of, and developmental precursor of, the PAR rules:

Rule A:

Seeing → Getting It Right (and Not Seeing → Getting It Wrong)

Rule A is the result of directly combining the two PAR rules used by preschoolers into a mechanism that could be used by nonverbal organisms such as chimps and human infants, and removing the proto-mental concept KNOW from the two PAR rules. Hedger and Fabricius (2011) conjecture that Rule A is operative in human infants and chimpanzees, and perhaps persists into adulthood.

More research into the anticipatory looking of children and adults would help to disconfirm or point in favor of the Rule A conjecture. However, the confound explained above first needs to be removed from the false belief task which studies to date have used. As discussed in Chapter 2 above, this can be done by using a two-task battery of a 2-location false belief task and a true belief task, as Fabricius et al (2010) demonstrated. Failing both tasks demonstrates use of RR, passing the false belief task but failing the true belief task demonstrates use of PAR, and passing both tasks demonstrates use of BR (see Chapter 2, Fabricius et al 2010, and Hedger & Fabricius 2011 for more details). Adults (and children) could be tested for Rule A by using the "eye gaze" methodology developed for children under 4 years of age (Clements & Perner, 1994; Garnham & Perner, 2001; Garnham & Ruffman, 2001; Ruffman, Garnham, Import, & Connelly, 2001) on both the true and false belief tasks.

Additionally, in order to know whether correct anticipatory looking in the false belief task indicates attribution of false beliefs or use of Rule A, the methodology needs to include a true belief task in which there is some interruption in the agent's connection to the situation that is comparable to what occurs in the false belief task; in other words, a *cue for situation change*. This can be done by simply having the protagonist leave and then return (see Chapter 2 and Hedger & Fabricius, 2011). Unfortunately, the previous eye gaze studies have not included such true belief tasks. If adults were to show anticipatory looking to the incorrect location on a true belief task despite giving the correct location verbally, then this would be a surprising finding that is nonetheless predicted by the Rule A hypothesis of Hedger and Fabricius (2011). Giving adult subjects a cognitive load or other working memory task may help prompt the use of a modular, implicit mechanism. New studies are therefore needed, and this is another project suggested by the present monograph.

As discussed in Chapter 2 above, if adults were found to continue to use Rule A implicitly it could explain a puzzling feature of ToM—that at times we appear able to make judgments about the mental states of others quickly, automatically, and effortlessly (Fodor, 1987) while at other times the process is difficult and deliberative. There is evidence that BR is effortful and difficult for adults (Apperly et al, 2006; Keysar et al, 2003). Lin, Keysar and Epley (2010) found that higher working memory capacity can have a positive impact on adult performance in theory of mind tasks, while cognitive load impairs this ability. Perhaps this can be explained by BR and Rule A being different psychological mechanisms, the former explicit and effortful and the latter implicit, automatic and modular.[128]

2.3 True belief task for infants

At this point there is deluge of studies on the false belief task in infants arguing for the *Nativist* view of ToM development, and many of these purport to include a true belief task as a control. However, as explained in Chapter 2 above and in Hedger and Fabricius (2011), to date these tasks have lacked a suitable *cue for situation change* that is comparable to those found on

---

[128] I address worries about the existence of both a verbal PAR mechanism (during the second stage of ToM development) and a modular Rule A mechanism (evident in infancy but perhaps persisting into adulthood) and related concerns in Chapter 2, Section 7, above.

false belief tasks. As Hedger and Fabricius (2011) note, the use of rules which predicted that a person wouldn't be able to find an object as soon as it was occluded from view would presumably be unadaptive (e.g. it would lead to unsuccessful caching of food and other items needed for survival). One would expect such a mechanism not to survive the natural selection process.

However (as discussed in Chapter 5 above), the survival of Rule A can be explained by the fact that it does cause successful predictions of behavior in limited situations, such as those similar to the 2-location false belief task. In fact, there is some evidence that Rule A (or a similar modular mechanism) can be found in chimpanzees (see Hedger & Fabricius, 2011). After all, even given an understanding of representational mental states, we can recognize that subjects who can't see an object will very often "get it wrong" and look in the incorrect location, as Rule A predicts.[129]

The reason for these successful predictions is that subjects compute the rule every time the protagonist is in a different situation, but if the protagonist remains in the current situation then subjects will not compute Rule A again. For instance, if I hide a Snickers candy bar underneath my *Spirited Away* DVD directly in front of Chantel while she reclines on our couch, even a baby using Rule A should not predict that Chantel will "get it wrong" and look in a different location, because this protagonist (Chantel in the current example) is still in the same

[129] I was reminded of this when I misplaced my cell phone this morning. Even given familiarity with the possible locations of the cell phone (viz. various locations in my apartment and my car), and extreme familiarity with the habits of the protagonist (since said protagonist was myself), I looked in a frustratingly large number of incorrect locations before finally stumbling upon my phone. (I can only hope that unlike Wittgenstein, my philosophical method is significantly better than this: "I do philosophy now like an old woman who is always mislaying something and having to look for it again; now her spectacles, now her keys," Wittgenstein, 1969, Section 532).

situation even after her perceptual access to the candy bar is broken (i.e. the Snickers is occluded from her view by the DVD case). We require an adequate cue for situation change (e.g. Chantel leaves the room and comes back) which will trigger the infant (or chimp or adult) to compute Rule A again for the current, new situation.

This of course explains why Rule A causes failure on true belief tasks. However, none of the true belief tasks found in the infant studies so far have such a cue (see the review in Hedger & Fabricius, 2011, for detailed individual explanations). Therefore, an infant true belief study that does contain such a cue is needed in order to test the Rule A hypothesis. Furthermore, a review of more recent infant studies is needed in order to demonstrate that such a test has not yet been done, and that brings me to the next project.

2.4 Further development of the argument against *Nativism*.

As we just mentioned at the end of the previous section, one project which would be useful is a continuation of the review of infant and chimpanzee studies from Hedger and Fabricius (2011) updated to include discussion of the studies done since 2011. Along with a review of empirical studies, with individual discussions of the true belief tasks used and (where appropriate) explanations of the lack of suitable cues for situation change, there have also been a few theoretical papers prompted by the current nativist euphoria[130] in developmental psychology (see e.g. Carruthers 2013, Southgate 2014). Thus it may be useful to also apply the theoretical argument against Nativism from Chapter 2 more specifically to these papers. One

---

[130] Hat tip to Nagel 1974.

challenging issue which needs to be addressed is getting clear about what it means for ToM or BR to be innate.[131]

2.5 Testing the A/R explanation

As I explained in Chapter 3, there is also a method we could use to test the explanation I give there for the success of PAR subjects on Appearance/ Reality (A/R) identity tasks. In the Standard A/R Task, Flavell et al (1983) showed children a sponge which looks like a rock, and then allowed them to handle it. After the child is familiarized with the fake rock, it is placed on the table. Children are then asked, "When you look at this with your eyes right now, what does it look like?" [Appearance Question] and "What is this really, really? Is this really, really a rock or really, really a sponge?" [Reality Question] (p. 102). 4- and 5-year-olds tend to get these questions correct, by responding to the appearance question with "a rock" and the reality question with "a sponge."

In Chapter 3 I presented the hypothesis that children using PAR interpret the appearance question not as "How are you visually representing the object?", because they lack the concept of mental representation. Instead they interpret the appearance question as "Do you have perceptual access to the object?" Hence this question triggers them to compute the

[131] Carefully distinguishing innate from acquired characteristics in psychology is notoriously difficult (see Griffiths, 2009, for discussion). Chomsky's account of language has been highly influential to contemporary discussions of nativism in psychology and the philosophy of psychology. His "poverty of the stimulus argument" is roughly that the small amount of exposure and feedback available to young children concerning grammaticality of language strings is insufficient to explain the vast knowledge of grammar exhibited by their linguistic behavior, including frequent production of novel grammatical strings (Chomsky, 1959). Here the relevant contrast seems to be between knowledge which is innate and knowledge which is the result of a cognitive learning process (cf. Fodor, 1975).

PAR rules. Thus a child using PAR should reason that I'm not touching the sponge *right now* (recall that the appearance test question includes the locution 'right now'), so I don't know it's a sponge, and I'll get it wrong. "Getting it wrong" in this case means thinking that it's a rock when it's really a sponge, and so the PAR subject should respond to the appearance question with "a rock." As with the False Belief Task and the A/R Property Task, the PAR child gives the correct response but means something very different by it than the typical adult does.

We could test this hypothesis by giving children a modified A/R identity task with three options, similar to the one used by Fabricius and Khalil (2003). First we familiarize the subject with a sponge that looks like a rock. We ask them what it looks like, then allow them to touch it, and so on. Then we show the child that a doll (Yoshi) rides on the rock-sponge as if it were a car, as Fabricius and Khalil (2003) did in their version. Now we are in a position to add a third option to the appearance test question: "When you look at this with your eyes right now, what does it look like? Does it look like a rock, or look like a sponge, or look like a car?" The present hypothesis predicts that 4- and 5-year-olds who pass the standard A/R task should divide their answers between "a rock" and "a car" but never answer with "a sponge."

Using PAR, the subject should reason that she isn't touching the object right now, so she doesn't know (Rule 1), and since she doesn't know, then she'll get it wrong (Rule 2). Some preliminary testing would help evaluate whether this is a suitable paradigm for a modified A/R task which would be adequate for testing my hypotheses concerning children's understanding of perception during the PAR stage, but at any rate this is another future project based upon the research program set out in this monograph.

3. Philosophical Problems

In this monograph I used the PAR hypothesis to shed light on two thorny problems in Cognitive Science and philosophy of psychology, viz. theories of concepts (Chapter 6) and what I call the *Perner-Povinelli Challenge* of designing an adequate empirical test for distinguishing between the use of reasoning involving genuine understanding of the mental and reasoning involving merely behavioral generalizations (Chapter 5). Of course, there are many other problems in the philosophy of psychology, and I believe the PAR theory can help us to find solutions for many of them. Following are brief descriptions of three of these issues, as well as suggestions for helpful applications of the PAR theory.

3.1 The Williamson view that knowledge is prior to belief

The standard view in philosophy is that *knowledge* is a composite, composed of the mental state of *belief* and other non-mental factors—viz. the truth of the belief, the believer having sufficient justification for believing the proposition, and some further conditions in order to avoid Gettier (1963)-type counterexamples.[132] There is such widespread agreement on this matter that few explicit arguments have been made for knowledge being a composite of which

[132] The failure of epistemologists in specifying suitable conditions which would avoid Gettier cases is one reason Williamson thinks the standard view is wrong (1995, 2000).

Some may object to the idea that a mental state is partly constituted by external factors such as truth, since in Williamson's view knowledge is purely a mental state (2000, pp. 21-48; see arguments to this effect in e.g.Fricker 2009, Magnus & Cohen 2003). However, those philosophers who already accept the externalism of Burge (1979) and Putnam (1975) shouldn't find this surprising or worrisome. At any rate, this issue cuts deeper than questions about knowledge.

belief is a simple part, but there seem to be at least two major reasons for the popularity of this account.

The first reason is due to the examples philosophers have used to explore the question, “What is knowledge?” In Plato’s *Meno*, for example, Socrates notes that a guide who knows the way to Larissa and a guide that merely has a true opinion about the way to Larissa will both be equally useful to the traveler, since both should be equally successful in getting her there. Yet knowledge is more highly prized than mere true opinion, as both Meno and Socrates note. Knowledge intuitively seems better and more valuable than merely correctly guessing for instance. The intuition is that both guides believe something true, but that something extra is required for someone to know something; the belief must be “tied down” like a Daedalus sculpture, and this tying down example suggests that to know is both to have a true belief *and something else* (which Socrates called an account or *logos*).

Another reason is that in discussions of folk psychology philosophers take belief to be central (see e.g. Churchland 1981, Davidson 1963, Dennett 1981, Fodor 1987, Horgan & Woodward 1985, Lewis 1972, Ramsey, Stich & Garon 1990). As mentioned previously, belief and desire are mental states that are the causes of human behavior, and typically if one knows what a protagonist believes and what a protagonist desires then one is able to predict and explain that protagonist’s behavior. For instance, if I know that Bob wants his Dretske book, and believes that his Dretske book is in his office, then I can predict that Bob is going to look for his Dretske book in his office (or explain why Bob is currently rummaging through the stacks of books piled about his office).

What's more, I can make this same prediction whether Bob knows his book in his office, falsely believes his book in his office, or has a "mere" true belief that his book is in his office (perhaps because his colleague Kevan is attempting to trick Bob, and believes the Dretske book is actually in his mailbox where he moved it yesterday, but unbeknownst to Kevan the secretary has noticed the book and returned it to his office). In other words, the generality of belief makes it appear that belief is more basic, and knowledge is just a special case of belief where some other non-mental factors such as truth and justification are also involved. However, the mental state of belief is all that is required to explain human action, and it is the thing actually causing the behavior.

A notable exception to this standard view is Timothy Williamson, who argues that knowledge is a simple mental state and is actually more basic than belief (1995, 2000). Part of the argument for this is that in some cases, knowing something instead of merely believing something can make a causal difference in human action. For instance, other things being equal, whether Bob knows his Dretske book is in his office or believes his Dretske in his office (even when the belief is true) can make a noticeable difference in how Bob behaves. Taking the earlier cases, when Bob knows his book is in his office he is likely to be more persistent in searching for it than in the case where he has a true belief because Kevan told him his book is in his office.[133]

Therefore there is reason to distinguish knowledge from belief in folk psychological explanations, because the two cases make a noticeable difference in behavior. Merely

---

[133] This example is of course inspired by Williamson (2000)'s case of the burglar on page 62.

mentioning Bob's belief that his Dretske book is in his office fails to sufficiently explain why Bob spends hours diligently going through all of his bookshelves, and even carefully going through the stacks of books piled about on his desk and on the floor, going through boxes etc. that Bob does when he *knows* his book is somewhere in that darn office. Thus knowledge is sometimes the appropriate *explanans* in folk psychological reasoning, and according to Williamson knowledge is also more basic than belief. The central and more common cases are the ones where someone knows something and that explains their behavior; cases of false belief and lucky guesses are parasitic upon the concept of knowledge (Nagel, 2013; Williamson, 2000).

Although not every belief is a case of knowledge, and hence knowledge remains a subset of belief on Williamsons's view as well, knowing is constitutively prior to believing and so the subset determines the nature of the larger set. Understanding believing requires understanding knowing in the sense that BELIEF is the derivative concept. Believing falsely is a kind of "botched knowing" (Williamson, 2000, p. 47) where knowledge was the aim but something went awry, and believing truly is a "diluted version of knowing" (Nagel, 2013, p. 285) where one lacks the credence of ordinary knowledge.[134] In folk psychological explanations, "believing *p* is, roughly, treating p as if one knew *p*, [so that] knowing is in that sense central to

[134] Another example of this is that knowledge is *factive*, meaning that Bob cannot know that his Dretske book is in his office unless it is true that his Dretske book is in his office. There are other factive verbs/ mental states, such as e.g. forgetting, remembering, acknowledging, regretting and so on, but Williamson (2000) notes that "knowing is important as the most general factive stative attitude" (p. 40), since these other factive states require knowledge in the first place. For instance, one can't remember X unless one knows X, etc. Here again, knowledge looks to be the primary, most basic concept among factives (Williamson 2000, pp. 33-41).

believing" (Williamson, 2000, p. 47). In all of these cases, one needs to understand knowledge in order to understand these variations upon that concept.[135]

If this picture is correct, and in particular if Williamson is right that understanding knowledge is necessary for understanding belief, then one would expect that the concept KNOW would develop before the concept BELIEVE.[136] In fact, Williamson takes this to be the case, and considers this another argument in favor his view: "A further ground for suspicion of analyses of the concept *knows* in terms of the concept *believes* is that they seem to imply that the latter concept is acquired before the former. Data on child development suggest, if anything, the reverse order (see Perner, 1991, pp. 145-203 for discussion of the relevant work)" (p. 33, footnote 7). Other philosophers have taken a more detailed look at the developmental data, but disagree on the interpretation. Nagel (2013) argues that the data do suggest that the concept of knowledge is acquired first, and so support Williamson's view, but Rose (2015) essentially argues that the data are inconclusive because results are conflicting. The PAR hypothesis has a good track record for making sense of apparently conflicting data in theory of mind development (see Chapters 2 and 3), so the theory may also prove useful in deciding this issue, and that is another future project I would like to do.

The PAR theory does in fact suggest that the concept KNOW is acquired before the concept BELIEVE, since the PAR stage makes use of the former but the latter is not acquired

[135] Williamson (2000) has other arguments for the primacy of knowledge, having to do with linguistic norms and knowledge's role in the speech act of assertion e.g., but I hope the foregoing discussion is sufficient for understanding his view. I focus on folk psychological explanations here as most relevant to theory of mind in developmental psychology. See McGlynn (2014) for a critical discussion of Williamson's view.

[136] If the concepts KNOW and BELIEVE are acquired at the same time, then that would also be consistent with Williamson's view as far as I can tell.

until the BR stage of development. However, one difficulty with this preliminary analysis is that although the PAR rules do reference whether a protagonist knows or doesn't know, this is a sort of proto-concept which differs from the full-fledged adult concept since children do not yet understand mental representation in the PAR stage. As discussed in Chapter 6, *knowing* for children in the PAR stage is a direct, unmediated connection to correct or successful behavior (and ignorance is a similar direct cause of unsuccessful behavior). Thus there is no such thing as guessing correctly but not knowing for the PAR child. Both guides to Larissa know according to the child using PAR, since both "get it right" by arriving in the correct location.

Although not the same conception of knowledge that adults and older children acquire later in the BR stage, the fact that children are using the word 'know' before they have acquired the concept BELIEF, and using (an admittedly simplified version of) the concept of knowledge in their reasoning during the PAR stage prior to acquiring BELIEVE, may nonetheless provide some support for Williamson's view. If KNOW was a compositional concept with BELIEVE as one of its parts, then it wouldn't be possible to have the concept KNOW before one had the concept BELIEVE, just as one couldn't have the concept BACHELOR before one had acquired the concept UNMARRIED. Arguably, the view of ToM development I advocate suggests that children at least begin to acquire the concept KNOW prior to the concept BELIEVE.

Williamson makes an intriguing comment when he notes that "children understand ignorance before they understand error" (p. 33, footnote 7), since use of the PAR rules involves attributing ignorance to the protagonist in the false belief task due to their lack of perceptual access. It might be argued that children master the concept of ignorance before they acquire

BELIEVE, since it doesn't depend as much on the idea of mental representation as KNOW and BELIEVE. What's more, as discussed in Chapter 6 children in the PAR stage also call a case of mistakenly telling a falsehood an act of *lying* (for instance, when accurately reporting the incorrect information to someone else because the protagonist was misinformed). This may show that children at the PAR stage already have some sensitivity to what Williamson (2000) calls the *knowledge norm of assertion*. These children seem to understand that a person should not assert what she does not know. At the very least, the developmental evidence is *consistent* with Williamson's view since children clearly do not acquire the concept BELIEVE before KNOW. As with all of these projects, a more in depth analysis may prove useful.

3.2 Fodor's compositionality constraint and my Theory-Theory view of concepts

In Chapter 6 I argued for a type of theory-theory view for certain mental state concepts based on the development of ToM. I argued that as children progress through stages of ToM development they use different concepts of mental states, and those concepts are at least partly individuated by the theories of mental states which children in those stages use (more specifically RR, PAR and BR). Although the theory-theory view is probably the most popular among developmental psychologists (Astington 1993; Carey 1985, 2010; Gopnik 1988, Gopnik & Meltzoff 1997; Gopnik & Wellman 1992, etc.), which is not surprising given the considerations I discuss in chapter 6, I don't know of any philosophers who advocate such a view. Fodor (1998) makes a number of important criticisms of the view (which also apply to most other theories about concepts, especially those held by psychologists). In Chapter 6 I addressed Fodor's shareability objection, but the other major objection which needs to be defended against is his compositionality objection.

To say that thoughts are compositional is to say that the meaning of a complex thought, such as e.g. the thought *Joe loves David*, is a product of the meaning of the parts of the thought and syntactical rules for the proper arrangement of those parts. Fodor believes that compositionality is the only way to explain three important features of thoughts.[137] First, thoughts are *productive*. This means that in principle (though not of course in practice) a person can think an infinite number of thoughts while having only a finite number of concepts. The thing which explains the productivity of thought is that thoughts are recursive, and this function is only possible if complex thoughts contain parts which contribute the same meaning each time they are used. The easiest way to see this is by using concepts such as *and*, *or*, etc. which allow one to add on to any thought. Hence, *Joe loves David* can become *Joe loves David and Chantel, Joe loves David and Chantel and he teaches philosophy*, and so on indefinitely. There are also for instance concepts represented by affixes in language which can likewise be repeated indefinitely: *I wonder what the world will be like for my great-grandchild, I wonder what the world will be like for my great-great-grandchild, I wonder what the world will be like for my great-great-great-grandchild,* etc.

The second feature of thoughts explained by compositionality is that thoughts are *systematic*. Systematicity means that if one can understand a thought, then one can understand a different thought containing the same semantic and syntactic constituents.

[137] In fact, it is these three features that are more important to Fodor than the claim about compositionality itself. See e.g. Fodor 1998, p. 94.

Common examples include, for instance, if a person can understand the thought *Joe loves Chantel,* then that same person can understand the thought *Chantel loves Joe*. That same person may not, however, be able to entertain the thought *Joe loves David* (e.g. when the thinker doesn't know the individual referred to by *David*). Similarly, the person who can think *Joe is taller than Chantel* can also think the thought *Chantel is taller than Joe* (but not necessarily the thought *Joe is taller than David*), etc. This is explained by the fact that both (1) although *Joe loves Chantel* and *Chantel loves Joe* are different thoughts, they both contain the same syntactic and semantic parts, and (2) if a person understands the meaning of those parts then that person can understand any thought with those same parts.

The third feature of thoughts that Fodor thinks is explained by compositionality is that different complex thoughts have identical parts, and those parts contribute the same exact meaning to those respective thoughts.[138] For instance, the component *brown* contributes exactly the same meaning in the thought *David is wearing a brown shirt* as it does in the thought *She has the most beautiful brown eyes*. Among other things, this context independence explains for example how definite descriptions are able to refer to an object without naming it (Fodor 1998, pp. 99-100). Thus one can understand and use the definite description *the brown purse* to refer to a (salient) object without knowing anything else about the object other than it has the properties of *being brown* and *being a purse*.

[138] Does this argument beg the question? If it does, then I assume it is my mistake and not Fodor's. This is my attempt to capture what Fodor (1998) calls "the best argument for compositionality," which is that "its traces are ubiquitous" (p. 99).

I agree with Fodor that (most) thoughts are compositional, and this best explains the productivity, systematicity, and other features of thoughts. The issue arises for my theory-theory view of mental state concepts because (assuming what I argue in Chapter 6 is correct) concepts such as KNOW are partly individuated by the overall theories of mind of which they are a part. For example, part of the content of the PAR child's concept KNOW involves psychological laws such as *knowing directly causes successful behavior* and so on—viz. at least the 2 PAR rules, and probably other psychological generalizations of the PAR theory as well. And the ToM of the PAR child differs from that of the adult. So the concept KNOW does not contribute the same meaning to the complex thought *Maxi knows where the chocolate is* for the PAR child and the adult.

One important thing to note from the start is that Fodor's general worry about holism (e.g. Fodor & Lepore 1992) doesn't apply here. This is the concern that since everyone probably has different beliefs, then if concepts are dependent upon a person's entire belief system no two individuals would mean the same thing by their concepts. If this were so, then thoughts couldn't be compositional since in effect the meaning of every concept (constituent of thought) would be relative to individuals. However, every child in the PAR stage shares the same ToM—or at least they share the basic components that determine the content of mental state concepts such as SEE and KNOW. The same is true for older children and adults using BR. Therefore the only difficulty arises between people using different theories of mind; e.g. when a child using PAR communicates with an adult. However, as I note in Chapter 6, it's an empirical question how well adults and PAR children communicate with each other about psychology, and my view is that there is a lot more talking past each other than is normally assumed.

As it stands, I don't see how this limited theory-theory is going to have a significant impact on the compositionality of thought. After all, even Fodor admits that there are non-compositional thoughts, such as idiomatic expressions for example. As Fodor (1998) remarks, to say that thoughts are normally compositional "means something like: *with not more than finitely many exceptions*. 'Idiomatic' expressions are allowed, but they mustn't be productive (p. 94; italics in original). Thoughts are productive and systematic in my view; even thoughts containing mental state concepts. For instance, whether the thinker is in the PAR or BR stage, if she can grasp the thought *Joe sees David* then she can entertain the thought *David sees Joe*. It's just that SEE means something different depending upon which stage of ToM development that thinker is currently in. In the BR stage SEE is representational but in the PAR stage it is not; so the two concepts are not the same. The concepts can be treated in the same way linguists treat semantic ambiguities. My view doesn't conflict with the compositionality of thoughts so far as I can see. The meaning of most concepts may still be atomic in the Fodorian sense.

Fodor's major worry about theory-theory views is his charge of inconsistency. He says that if concepts are partly constituted by the theories that contain them, then it is not possible to learn a new concept in place of the old one (1998; pp.115-119; cf. 1975). The meaning holism of theory-theory is incompatible with the view of cognitive development as conceptual change, according to Fodor. For instance, since the meaning of KNOW is partly constituted by the PAR theory for the 5-year-old, how do they learn the representational concept KNOW which replaces the outdated concept as they acquire the BR theory of mental states? The problem, according to Fodor (1998), is that "a theory can be used to effect the implicit definition of a new term only if at least some of its vocabulary is *isolated* from meaning changes of the sorts that holists say that concept introduction brings about" (p. 115; italics in the original). The first PAR rule, for example, involves a relationship between perception and knowledge, but the PAR concepts SEE and KNOW are constituted by the PAR theory itself. So I owe an explanation of how the BR child acquires the adult representational concepts SEE and KNOW, since the difference between these two stages of development is precisely the acquisition of a new ToM, which will in turn individuate new mental state concepts.

I agree that I do owe such an explanation, and another future project will be spelling out the details of this process, ideally coupled with more empirical research into the transition between the three ToM stages. However, I disagree with Fodor's *a priori* claim that this cannot be done. For instance, the PAR concept SEE will be related to other non-representational concepts, such as LOOK, DIRECTION, DISTANCE, OBJECT, ENVIRONMENT, and so on (see Chapter 3 above). These concepts are, in Fodor's terminology, isolated from the vocabulary which is dependent upon the PAR theory. Concepts such as these mean the same in all three ToM stages and may be involved in the transitions between them. Likewise, the PAR concept KNOW will be related to non-representational concepts such as SUCCESS, FAILURE, ACTION and so forth (see Chapters 2 and 6 above). It will take some work to cash out the details of this story, but I'm not concerned about Fodor's criticism being a death knell.

Granted, the most challenging part of the story of transition from PAR to BR will be explaining the acquisition of the concept of MENTAL REPRESENTATION, which is admittedly key to the mental state concepts contained in (and partly individuated by) the BR theory. How is this completely new conception of the mind acquired? However, here I have a hunch that the way around Fodor's criticism is to say that the BR theory is not *learned*. As sketched in Chapter 6, my hunch is that the BR theory (along with RR, PAR and Rule A) is *innate.* The reasons for this are: (1) these theories are universally found among all human beings regardless of their culture, upbringing or education, (2) these theories do not appear to be the result of any learning process, and (3) the sole major factor in the acquisition of these theories seems to be the *age* of the subject (e.g. Wellman et al, 2001).

The details of the argument can be found in Chapter 6 above, where I also note that Fodor (1987) already agrees with me that BR (what he would've called *folk psychology*) is innate. Fodor makes a case for this which could be added to my argument, including the lack of a rival explanation, and the fact that the complexity of the theory would likely make learning it too long a process for an individual life span. However, once again the details of this nativist theory need to be worked out, and experiments need to be designed to further test this hypothesis. (The most pressing challenge, which I shudder to think about, may be to give a fuller explanation of what exactly *innate* means.)

3.3 Further development of the implications for the HOT theory of consciousness

The PAR hypothesis presented in this monograph would seem to have drastic consequences for the Higher Order Thought (HOT) theory of consciousness. According to the HOT theory of consciousness, in order for a mental state M to be conscious, the subject must have a higher order thought (HOT) *about* that state (Carruthers 2000, Gennaro 2004, Rosenthal 1986). For example, take Armstrong's (1981) much discussed example of the long distance driver. Often when driving familiar routes or when travelling long distances on the freeway, we have an experience of "zoning out," where we are thinking about philosophy or listening to music or involved in conversation with a passenger. We may suddenly become aware of our surroundings ("Oh wow, look, we're already almost to Watertown," we think when we see the sign on 81 North) and realize we have been driving on automatic pilot but consciously concentrating on something else.

According to Armstrong, when this happens you were perceiving the road (and other vehicles, traffic lights, etc.) this entire time, but weren't consciously aware of this perception because you were consciously aware of e.g. *thinking about the hard problem of consciousness*.[139] You must have been perceiving the road, or else you would have crashed during that large turn 10 miles back. However, you weren't consciously perceiving the road, because you were aware of something else. Now that you have "come to" and are aware of driving once again, your perception of the road is conscious. According to HOT theory, what distinguishes the current, conscious perception from the former, unconscious perception is that there is a higher order thought directed at this perception of the road.

Details of what this HOT is supposed to be are lacking in the literature. At the very least, however, in order to be a HOT at all it must be *about* the perception of the road. Since concepts are the constituents of thoughts, then the HOT that makes the perception of the road a conscious perception must have the concept SEE (or the like) as one of its parts. However, according to PAR theory, the representational concept SEE is not acquired until the BR stage of

---

[139] Of course, Armstrong uses this example to motivate his own Higher Order Perception theory. However, HOP doesn't have the same obvious difficulty that HOT does, since a child is able to perceive before she is able to understand perception.

ToM development, or until about 6 years of age. Hence a consequence of the HOT theory would be that children under the age of six cannot have conscious mental states.

This problem with HOT theory, which Van Gulick (2006, p. 13) dubs the “too fancy” objection, has been discussed previously (Dretske 1995, Tye 1995). However, now the objection can be made more precisely, and there is clear empirical evidence in favor of it. Now we can say exactly what it is that is too sophisticated for a young child to possess, viz. an understanding of mental representation. Furthermore, the objection is also considerably strengthened. Previously, the “too fancy” objection was thought by many to apply only to infants, and not to school age children (cf. e.g. Gennaro 2012). Although I find the idea that infants don’t have phenomenal experience highly implausible, there have been philosophers who have notoriously bitten the bullet on this point (Carruthers 2000, 2005). However, it’s even less plausible to deny that 5-year-olds have conscious mental states. HOT theorists such as Gennaro (2012; Chapter 7) and Carruthers (2013) were hoping to be saved by the Nativist view (see Chapter 2 above), but the implausibility of ToM Nativism makes a strong case against the HOT theory of consciousness. To deny consciousness to a 5-year-old is preposterous.[140] Therefore, this makes a third useful philosophical application of the PAR theory to be written up in more detail in the

---

[140] The state of the art among HOT theorists is to embrace Nativism (Carruthers 2013, Gennaro 2012) , even for hardline theorists such as Carruthers who was at one time willing to deny that infants were conscious (2000). So the fall of Nativism, and a more protracted development than the Traditional View suggests, is a crushing blow for these philosophers. However, it should be noted that other HOT theorists have attempted to get around the “too fancy objection” by suggesting that HOTs do not require sophisticated concepts (see e.g. Rosenthal 1993). This may be a way around the issue I raise here for HOT theory, but then the HOT theorist owes an account of exactly what kind of non-conceptual thought a HOT is supposed to be, and how it is able to do the kind of work that the theory requires.

future. I also believe the theory can lend insight into what this phenomenal experience of 5-year-olds must be like; this is the subject of Section 4.

4. What it's like to be a 5-Year-Old

In terms of my longer term research project exploring major philosophical issues through the lens of the PAR hypothesis, I plan to use it next to examine issues of *consciousness*. In particular, I believe the PAR hypothesis, along with an understanding of the child's development of the concept of a self, can bring insight into the phenomenal experience of young children.

4.1 Theories of consciousness

There is much work in philosophy attempting to provide necessary and sufficient conditions for consciousness. The prospects for this sort of project, however, appear to be bleak. For one thing, it's difficult (perhaps impossible) to come up with a strict definition for almost any naturally occurring word (Fodor et al 1980). Also, this sort of program has a long history in Western philosophy—going back most famously to Plato but perhaps earlier—with an embarrassing track record. Outside of Logic, philosophy has failed to produce a single uncontroversial analysis of this sort after almost 3000 years. If we can't even give necessary and sufficient conditions for something as ordinary and unmysterious as a chair or a book, then it appears to be a waste of time attempting to come up with them for something as complicated and opaque as knowledge or consciousness.[141]

Just as Fodor (1975) provides a model for the methodology of this monograph (see Chapter 1), Fodor (1983) provides a model for how the sort of project I am undertaking here might take shape. Fodor (1983) does not attempt to give necessary and sufficient conditions for modular psychological mechanisms, but he notes that modular input systems form a natural psychological kind, and then goes on to discuss nine features which many of these systems share and which seem to demarcate a natural category. He argues "that the functionally specified class *input system* does pick out a 'natural kind' for purposes of psychological theory construction; that there are, in fact, lots of interesting things to say about the common properties of the mechanisms that mediate input analysis" (p. 44).[142] The common properties which Fodor elaborates on in Part III are neither necessary nor sufficient. He says that "there appears to be a cluster of properties that they have in common but which, *qua* input analyzers, they might perfectly well not have shared" (p. 101). Nonetheless, *qua* modules these properties are important for making them the kind of thing that they are.

Although not exactly analogous, this project will follow a similar model of theory-building.[143] By consciousness I shall be referring to what Block (1995) calls *phenomenal*

---

[141] Similar sentiments appear e.g. in Block (1995) and Searle (1992). Other philosophers argue that all (or many) concepts have fuzzy boundaries and lack strict necessary and sufficient conditions, most famously Wittgenstein (1953) but also Putnam's "cluster concepts" (Putnam 1962).

[142] Fodor characterizes a *natural kind* as "a class of phenomena that have many scientifically interesting properties in common over and above whatever properties define the class" (p. 46).

[143] Differences between my project and Fodor's include the fact that Fodor was taking a natural kind (input analyzers) and arguing that they share features which make them part of a distinct category, viz. modules. Another difference is that Fodor was able to delimit the category "input analyzers" because they form a *functional* kind, so

consciousness—the experiential mental states where there is *something it is like* (Nagel 1974) to have them.[144][145] I will take it to be obvious that the phenomenal consciousness of psycho-typical human adults forms a natural kind. As such, it has scientifically and philosophically interesting features. Van Gulick (2011), Part 4, outlines seven such features, and notes that the list is not exhaustive (p.25). One of these features is that phenomenal experience includes some awareness of the self (however minimal), which Van Gulick (2006) calls "reflexive self-awareness." Block (1995) for instance notes that "P[henomenal]-Conscious states often seem to have a 'me-ishness' about them, the phenomenal content often represents the state as a state of me" (p. 178). Levine (2001) calls this aspect "the subjectivity of conscious experience." It may even include some awareness of mental states, as for instance Kriegel (2003, 2009) argues. I believe this feature of consciousness can provide insight into the phenomenal experience of children in the PAR stage of cognitive development.

4.2 Outline of the project

If awareness of self and (perhaps) awareness of mental states are indeed important aspects of adult phenomenal experience, as I will argue, then the limited understanding of the

---

that he could give an account of mental architecture modeled on the biological taxonomy of organs. However, the function of consciousness is a much disputed matter in philosophy of mind.

[144] Block (1995) says, "P[henomenal]-conscious states are experiential states, that is, a state is P-conscious if it has experiential properties. The totality of the experiential properties of a state are 'what it is like' to have it… we have P-conscious states when we see, hear, smell, taste, and have pains. P-conscious properties include the experiential properties of sensations, feelings, and perceptions, but I would also include thoughts, wants, and emotions" (p.166).

[145] Kriegel (2009) complains about using the "what it's like" locution to describe phenomenal consciousness, because different philosophers disagree about which states are captured by this phrase (p. 4). However, this latter controversy is tangential to the present project. Any states which have this character (recognizing that different philosophers will disagree about which states do in some cases) are what I am referring to by 'consciousness' in this project. Kriegel (2009)'s own rigidified definite description is too vague and uninformative as it stands, and too narrow for present purposes whenever precisified.

self and mental states which young children exhibit means that either children lack phenomenal consciousness, or else their phenomenal experience differs from that of psycho-typical adults in some interesting ways. Since I take the first option to be obviously false, this project will explore what it is like to be a young child, and how the phenomenal experience of young children differs from that of adults in some interesting and important ways.[146]

Thus, we begin with the idea that an essential aspect of phenomenal experience is the *self*, or (better) subject of consciousness,[147] which is experienced in a few ways. First, all conscious experience converges on a single subjective point of view. Second, practical day-to-day maneuvering through an experienced environment requires understanding ourselves as agents. Third, understanding our phenomenal mental states likewise seems to require a sense of ownership, that these experiences are *my* experiences. I am the author of my internal monologue. Some support for this is that a breakdown in this sense of ownership may result in psychological disorders. Perhaps e.g. a lack of this leads to the intrusive thoughts experienced by schizophrenics or persons with OCD (American Psychiatric Association, 2013). Finally, the narrative self of our episodic memory also ties together memories of my past experiences with plans for the future. Without some subject of consciousness who experiences all of these

[146] One question to be considered is whether consciousness goes through stages in a fashion similar to ToM development (and perhaps even because of it), or whether consciousness should be thought of as a continuum or spectrum in which typical adult and child consciousness appear at different points.
[147] My reasons for avoiding talking about the *self* involve the vagueness of the concept and that it often has an interpretation which is much thicker than required, and may push us into a version of what Van Gulick (2006) dubs the "too fancy" objection. (This would be the idea that creatures who aren't sophisticated enough to have a self intuitively have phenomenal experience. I'm doubtful whether cats e.g. have a robust sense of self, but I take it as obvious that they have consciousness.) Furthermore, I think it's an open empirical and philosophical question whether there is such a thing as a self in the robust sense. The state of the art in psychology suggests that we instead have "a loose confederation of sub-selves" (Block 2005; see also Gazzaniga 1998, Kurzban & Aktipis 2007). Lastly, philosophical discussions of the self are often concerned with *personal identity*, which is a metaphysical issue and tangential to the current project.

mental states, episodic or autobiographical memory wouldn't be possible.[148] Likewise, deficits in episodic memory can create a loss of identity (e.g. the case of William in Sacks 1985, Chapter 12).

Since they lack a conception of mental representation before the age of six (a central thesis of the present monograph), young children are not aware of their mental states *as mental states.* They don't have the sense of ownership or sense of agency which is so crucial to the phenomenal experience of adults. As the studies on Level 1 and Level 2 Perspective-Taking demonstrate (see Chapter 3), they are also not aware (at least at an early age) that different people experience the world from different subjective points of view. Arguably, then, they may not even realize that their phenomenal experiences converge upon a single, subjective point of view. Since young children lack a conception of mental states, they also lack a conception of themselves as conscious subjects of mental states. Thus, their phenomenal experience is fundamentally different from that of psycho-typical adults. So runs the argument of this project in broad outline.

One reason for thinking that these features are important is that deficits of these aspects arguably characterize at least some aspects of psychological disorders such as schizophrenia and identity disorders (American Psychiatric Association, 2013). There is already some precedent for what would be characterized as a psychological disorder in adults to be normal during the course of childhood development. For instance, obsessive-compulsive

---

[148] One interesting question is whether episodic memory and autobiographical memory are the same thing. Gopnik (2009) makes an interesting suggestion that young children may have the former, because they can remember specific events, but lack the latter, because they don't organize those memories into a coherent timeline (p. 147).

behaviors which would be considered pathological for adults are prevalent during the toddler years of normally developing children (Evans et al 1997). Although I'm not suggesting that identity disorders or the like are a normal stage of cognitive development, it is possible that normal development involves some related defecits; e.g. it may be normal to lack a coherent stream of consciousness or episodic memory for a time. For instance, it is known that children under the age of five don't tell the type of narratives normally associated with episodic memory, and don't verbally demonstrate any sense of ownership of their memories (Nelson & Fivush 2004). They also have difficulty freely recalling their memories without being prompted (Ornstein et al 2006).

4.3 Chapter plan for the project

After an introduction and overview of the project in Chapter One, Chapter Two ("Self-Reflexive Consciousness") will begin with some intuition pumps that typical adult phenomenal experience includes some understanding of the self, and will discuss Van Gulick's sketch of reflexive self-awareness as a starting point for the positive descriptive task of outlining the self-reflexive aspect of consciousness. Next I will examine Kriegel's view, which is in its most basic form the thesis that a mental state is conscious iff the subject is aware of it, and that our awareness is constituted by the self-representation of phenomenal states. I will argue that this thesis admits of a *de re*/*de dicto* ambiguity. Roughly, Kriegel leaves it open whether subjects are aware of mental states *in some way or another (de re)* or *as such (de dicto)*.[149] The same

[149] Rosenthal's HOT view admits of the same ambiguity. Making this distinction allows a sharpening of Dretske (1993)'s criticism of Rosenthal, but at any rate the challenges for HOT-theory enumerated in Van Gulick (2006) and Kriegel (2009) make the view extremely problematic in my view.

ambiguity applies to representation. The *de re* reading is true but trivial, and the *de dicto* reading is false because children are conscious but lack awareness of mental states as such until about six and a half years of age.

Chapter Three ("The Subject of Consciousness") will draw on various sources from analytic philosophy, phenomenology, and psychology to present an original and detailed analysis of the self-reflexivity of phenomenal consciousness. Part of this project will involve distinguishing between various theses which sometimes go under this umbrella. There seem to be four important aspects of the self-reflexivity of consciousness (outlined in the previous section) which are conceptually distinct but certainly not unrelated.

First, all aspects of our phenomenal experience converge upon a single point of view. Consider for instance our phenomenal experience as we sit in a tavern. The cold glass of beer in front of me, the bartender moving behind my focus, the people getting less and less vivid as they are arranged further into the periphery, the smells of pizza and cigarette smoke wafting in from different sides of me, the humming noise of various indistinguishable conversations coming from different directions in the background as I focus on what my friend is telling me directly to my right, the feel of the hard wood on my back which I only become aware of as I focus on it, the football game being played on a speaker above me and to the left, the cold wet handle of the glass mug and the weightiness of the contents as I hoist it to my mouth, the crisp refreshing taste of the beer as I drink, and the proprioceptive awareness of my body in space. At any one time, we have an incredibly rich and varied array of experiences which are directed at our own unique perspective upon the world. There is a vast sea of experience, what Searle

(2005 and elsewhere) calls the "unified conscious field," which always converges upon a single vantage point, with an obvious area of focus and other experiences fading further and further into the background from that focus, until—like the feel of the shirt on one's body and familiar background noises—we seem to lose awareness of them altogether at times. Let's call this the *subjectivity* of consciousness.[150]

It's difficult to conceive of phenomenal experience which lacked this subjective point of view, but it's hard to tell whether that is of metaphysical import or simply a limit of our imagination. Films, for example, convey images coming from various points of view rather than a single one (think of how jarring and unsettling it would be if our normal experience was like this!), but nonetheless all of the images have *some* point of view. Supposing God exists, perhaps his phenomenal experience would be of the totality of the world from a third-person perspective (something like the "God's-eye point of view" of Gabriel García Márquez's novels). However it seems impossible to conceive of what this would be like. Some philosophers, including Searle (2005), Kriegel (2009), and Sartre (1943), contend that consciousness without this subjective character is a metaphysical impossibility.

Second, there is a sense of agency accompanying our phenomenal experience and our practical day to day getting around in the world. We feel ourselves to be the causes of our actions, and this seems to be necessary in order to do the kinds of planning and making rational sense of human behavior that we do. To begin with, consider the sea of experiences at a tavern

[150] Perhaps it should be noted that it's not at all clear whether this is what Kriegel (2009) means by "the subjective character of consciousness." At any rate, Kriegel's characterization of the latter as the "for-me-ness" of our phenomenal consciousness is much too vague to do any real work in the present context.

that we used as an example two paragraphs back. Accompanying all of these experiences there is also the sense that I have some voluntary control over them. I can change my focus, for example. I could stop attending to my friend's conversation and listen to the ball game, or turn my visual attention from the mug of beer to the bartender making drinks. I can also change the conscious field by simply turning my head. We can imagine a being who experiences everything passively, who is unable to turn her head or move her eyes, and whose focus of attention is constantly being pulled from one thing to another without her control, in the way that our attention is sometimes "grabbed"—for example in the cocktail party effect when I hear my name coming clearly from the babble of background conversations behind me. This kind of experience is conceivable, but it would be very different from our own.

Beyond this, we could also imagine a being who had no control over her decisions and actions, whose consciousness inhabited a sort of robot which was completely under someone else's control. When a button is pushed, she feels the arm of the robot raise. However, our own phenomenal experience includes a robust sense of freedom of the will (even if philosophical argument convinces us that this is an illusion). Actions don't just happen, but rather I decide and act. If this weren't the case, then interpreting human behavior in terms of reasons for acting and responsibility for certain actions would not make sense. The latter practices would also seem to depend upon the subjectivity of consciousness. In order to make sense of the idea that my thirst gives me a reason to drink beer, it must be the same subject who sees the beer, feels the thirst, makes the conscious decision to drink, and then lifts the glass to the lips and all the rest of it. The sense of agency is related to what Van Gulick (2006) calls "the pragmatic aspect of our experiential content"—things are present to us in experience as possibilities for practical engagement with the world.

Third, there is what I call *the sense of ownership* we have over our phenomenal states. An important aspect of our phenomenal experience is that these are *my* experiences. I am the one smelling the pizza and hearing the football game. I need to understand that I am the one seeing the mug and desiring the beer in order to make sense of my decision to lift it and drink it. I need to understand that I am the one seeing the truck coming towards me in order to jump out of the way. People sometimes lose this sense of ownership over their own body parts, and feel for example that their own leg belongs to someone else (Vallar & Ronchi 2009). This phenomenon must include a corresponding loss in proprioception. It's not clear how often similar deficits could occur for mental states, but a plausible explanation for the intrusive thoughts experienced by schizophrenics is that they lack this sense of ownership for some of their internal monologue.

Fourth, there is our episodic memory, our autobiographical narrative which connects past experiences with future plans. Psychologist Endel Tulving (1985) distinguished semantic memories, such as remembering that Albany is the capital of New York, from episodic memory, such as remembering your first kiss. Episodic memories include recollection of phenomenal experiences, such as remembering the feel of rain on one's face, the touch of soft lips, the smell of perfume, and the sense of embarrassment one felt after realizing that the kiss was being watched by a teacher. That phenomenology has the same subjectivity and sense of ownership that occurrent phenomenal experiences exhibit. Episodic memory also allows us to organize past events coherently and make plans for the future based upon past events.

This sort of memory is importantly related to constructing a sense of self which persists through time. Persons with dissociative identity disorder have distinct timelines and memories for their separate personalities.[151] When preschool and kindergarten children are asked to recall their day, they don't tell the same sort of story full of self-references that adults would ("I was talking with Santos and the teacher told me to be quiet" etc.) but instead give what psychologists and computer programmers call "scripts" ("Well, first the teacher gets everybody together in a circle, and then you sing, and then it's naptime, and after that you paint" etc.) [Fivush 2011]. There is some important connection between self-awareness and episodic memory.

Lastly, although it seems at least conceptually possible that any of these four features could exist without the other three, the occurrence of all of them together would appear necessary in order to give the sense of unity and coherence which our phenomenal experience exhibits. Also, there may be some important dependency relations between them. For instance, understanding ourselves as agents may require the subjectivity of consciousness and the understanding that these are *my* perceptions, motivations and decisions (i.e. a sense of ownership). Perhaps a sense of agency is required for a sense of ownership. Here I am thinking of certain anosognosia patients who are unaware of their inability to move a body limb, and

[151] For instance, Baars et al (2003) note that "patients with identity disorders such as fugue (a rapid change in personal identity lasting weeks or months) often show amnesia for the eclipsed self. When the patient returns to normal, he or she might report time loss—a period of weeks from which no conscious experiences can be recalled. It as is if each personality serves to organize and interpret conscious events during its time of dominance" (p.673).

report thinking that it belongs to someone else. At any rate features two and three seem tightly interconnected. Also, a sense of ownership appears necessary to have a coherent autobiographical memory. If I don't recognize that I was the person receiving the email last month, then I have no reason to revise and resubmit the paper by the deadline next Friday.

Chapter Four will examine the level of understanding of the self as a subject of consciousness and the level of understanding of mental states which is entailed by the analysis of the previous chapter. In other words, how much must a subject understand about mental states and the self in order to have this sort of phenomenal experience, which is the kind psycho-typical adults have? This is a difficult issue because, on the one hand, robust intellectual understanding of the self and mental states is obviously not necessary. After all, there is still plenty no one knows about mental states and the self. On the other hand, *some* awareness of the self seems to be necessary for psycho-typical adult consciousness.

Chapter Five ("Seeing and Believing: Children's Understanding of Mental States") will summarize the evidence presented in this monograph, demonstrating that children under the age of roughly 6.5 years lack a conception of mental representation. This monograph explored children's understanding of belief (Chapter 2), perception (Chapter 3), and mental concepts in general (see especially Chapters 5 and 6). If some of the work outlined in Section 2 (exploring children's understanding of desire and other mental states) above is completed by this point, this will also be summarized in Chapter 5 of *What it's like to be a 5-year-old*. The cognitive development of theory of mind sheds light on the consciousness of children in the early stages of development, viz. the reality reason (RR) and perceptual access reasoning (PAR) stages. For instance, understanding perception involves understanding that different perspectives yield different visual representations. As I explain in Chapter 3, Flavell (1974) distinguishes two kinds of perspective understanding.

Consider Level 1 perspective taking. To understand seeing, one must be able to recognize that different people see different things. This is a very rudimentary, basic belief, which follows straightforwardly from recognizing that things look differently from different perspectives, combined with the recognition that it must also be like this for others. Understanding visual representation minimally requires recognizing that different people can see different things than what I see. Surely, someone who answers that she sees a dog (when the picture is facing her but away from the experimenter), and that the experimenter also sees a dog, doesn't understand what it means to see. Yet, Moll and Tomasello (2006) found that 18-month-olds gave the correct answer in a version of a Level 1 task less than half of the time (i.e. they were performing at less than chance). Thus these young children lack even a rudimentary understanding of perceptual states, and their perceptual experience probably does not include a sense of themselves as agents, and maybe not even a realization that their perceptual experiences converge upon a single, subjective point of view.[152]

Mental states such as perceptions are of course *representations.* Furthermore, as philosophers initially suggested to psychologists (Bennett, 1978; Dennett, 1978; Harman, 1978), a good way to test whether someone understands mental representation is to test whether they are sensitive to misrepresentation. It would be implausible to say that someone can

---

[152] See Chapter 3 above for more on children's understanding of perception and its development.

understand representation without realizing that representations can sometimes *misrepresent* reality. Flavell and colleagues have also explored when children learn that reality can be different than it appears (the appearance/reality distinction; Flavell et al, 1983; see Chapter 3 above). One task they used to test this involved showing children a sponge which looks like a rock, and then allowing them to handle it. After the child is familiarized with the fake rock, it is placed on the table. Children are then asked, "Is this really a rock or really a sponge?" and "When you look at this with your eyes right now, does it look like a rock or look like a sponge?" They found that before four years of age children tend to give the same answer for both questions (i.e. they say that it really is a sponge, but also that it looks like a sponge), presumably because they don't realize that things can be different than they appear to be.[153] Flavell, Green & Flavell (1995) also found that 5-year-olds failed to understand *thinking about something*, even after familiarization and education. As explained in Chapter 2 above, probably the most startling lack of understanding of mental representation by children as old as four and five is evidenced by their failure on the true belief task (Fabricius et al, 2010).

This finding is significant, because part of what is required to pass this true belief task is the understanding that Maxi has a representation (belief) of the location of the chocolate which will persist through a brief departure and subsequent return. When combined with other findings in theory of mind development, this makes a compelling case that children younger than 6 years of age do not understand mental representation, as this monograph argues. Furthermore, lacking this sort of understanding would seem to preclude them from

[153] In my view, 4-year-olds don't understand mental representation, so there is no such thing as appearances for them; there is only how the world is, and access to that world or lack thereof. See Chapter 3 for details.

understanding themselves as owners of such states, and experiencing the world as the sort of interplay between agents and things presented to consciousness that Van Gulick (2004, 2006) describes. Thus, their conscious experience is importantly different from our own.

Chapter Six ("Understanding of the Self in Young Children") will examine the relevant empirical literature from developmental psychology, use the analysis in Chapter Three to organize the findings, and argue for a view concerning children's understanding of the self. When children acquire a sense of self is an issue of current debate amongst psychologists and philosophers, so getting clear about this will be difficult but worthwhile even outside of the current project. Some argue that the work of Meltzoff and Moore, which demonstrates that newborns are capable of imitating facial expressions of an adult experimenter, is sufficient to demonstrate that an awareness of the self is innate (Gallagher & Meltzoff 1996, Wider 2006).

However, children don't pass the standard mirror test until about eighteen months to two years of age (Brooks-Gunn & Lewis 1984),[154] and Daniel Povinelli has found that adding a small amount of time (about 5 minutes) to the mirror test will cause 3-year-olds to fail (Povinelli et al 1999, Povinelli 2001). He and his colleagues surreptitiously placed a sticker on children's heads while they were being videotaped, and then immediately watched the video alongside the young subjects. 5-year-olds reached for the sticker on their own foreheads, but younger children did not, even though they recognized that the person in the video was them. Povinelli

[154] This is the well-known test where a sticker is placed on the baby's forehead and then she is shown her own image in a mirror. Touching the sticker on one's head rather than the one in the mirror constitutes passing. (Incidentally, I'm not sure how much stock I put in this test. Cats, for instance, are not supposed to pass; but I once witnessed my own cat turn from staring at his image in the mirror to swat at a fly which he first saw in the mirror. He must have realized that he was looking at himself in the mirror in order to do this. I suppose the mirror test is sufficient but not necessary for some understanding of self.)

argues that 3-year-olds don't connect the past self with the current self, because they didn't realize that putting a sticker on them five minutes ago meant that the sticker would still be on their forehead.

What is required for the sort of self-awareness involved in adult phenomenal experience requires not only synchronic recognition of the self in the moment, but also an understanding of the self through time.[155] Hence understanding children's episodic memory will also be important, and there is some evidence that young children don't have the concept of a diachronic self. Even after they are able to remember specific past events (episodic memory), young children have difficulty putting those events in order in a timeline. For instance, 3- and 4-year-olds—who are very accurate at telling which pictures they have seen before and which they haven't—have difficulty telling whether they saw the picture today or yesterday (McCormack & Hoerl 2005, 2007).

Finally, Chapter Seven ("The Phenomenal Experience of Young Children") will comprise an informed speculation about what the phenomenal experience of children at various stages from birth to about six and a half years of age must be like in comparison to that of the psycho-typical adult, given their limited understanding of the self and of mental states during those stages. One obvious constraint will be that the view is compatible with the developmental psychology studies of consciousness in children (Flavell et al 1993, 1995, 1997, 1999, 2000);

---

[155] There is an interesting parallel with the theory of mind literature here. Many researchers attributed an understanding of mental states too early in development because they failed to notice that understanding mental representations requires understanding that they are maintained through time in a person's mind, and are not merely synchronic episodes connected to the current situation.

however there isn't much of the latter to go on.[156] Finally, I will make a comparison between my (eventual) view and that recently put forward by Alison Gopnik (2009, Chapters 4 and 5).[157]

[156] Also, this literature might turn out to be irrelevant insofar as it focuses on children's *understanding* of consciousness rather than their actual phenomenal experience.

[157] I would like to acknowledge the following for useful discussions about some of the ideas presented in this chapter: Bill Fabricius (Sections 2.2, 2.3 and 2.4), Bob Van Gulick, Bernie Kobes (Section 3.1), Kevan Edwards (Section 3.2), and Ty Boyer (Section 2.2).

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Anscombe, G. E. M. (1957). *Intention*. Cambridge: Harvard University Press.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17,* 841-844.

Armstrong, D. M. (1981). *The nature of mind and other essays.* Ithaca, NY: Cornell University Press.

Astington, J. W. (1993). *The child's discovery of the mind.* Cambridge, MA: Harvard University Press.

Baars, B. J., Ramsey, T. Z., & Laureys S. (2003). Brain, conscious experience and the observing self. *Trends in Neuroscience, 26,* 671-675.

Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences, 1,* 557-560.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18,* 227-247.

Block, N. (2005). Interview. In S. Blackmore, *Conversations on consciousness*, pp. 24-35. Oxford, UK: Oxford University Press.

Brooks-Gunn, J. & Lewis, M. (1984). The development of early visual self-recognition. *Developmental Review, 4,* 215-239.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Carey, S. (2010). *The origin of concepts.* Oxford, England: Oxford University Press.

Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory.* Cambridge, England: Cambridge University Press.

Carruthers, P. (2005). *Consciousness: Essays from a higher-order perspective.* Oxford, England: Clarendon Press.

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

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal behavior*. *Language, 35*, 26-58.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*, 67-90.

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

Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy, 60,* 685-700.

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences, 1*, 568-570.

Dennett, D. C. (1981). True believers: The intentional strategy and why it works. In A. F. Heath (Ed.), *Scientific explanation: Papers based on Herbert Spencer lectures given in the University of Oxford.* Oxford: Clarendon Press, pp. 150-167.

Dretske, F. (1993). Conscious experience. *Mind, 102,* pp. 263-283.

Dretske, F. (1995). *Naturalizing the mind.* Cambridge, MA: MIT Press.

Evans, D. W., Leckman, J. F., Carter, A., Reznick, J. S., Henshaw, D., King, R. A., & Pauls, D. (1997). Rituals, habits, and perfectionism: The prevalence of and development of compulsive-like behaviors in normal young children. *Child Development, 68,* 58-68.

Fabricius, W. V., Boyer, T., Weimer, A.A., & Carroll, K. (2010). True or False: Do five-year-olds understand belief? *Developmental Psychology*, *46*, 1402-1416.

Fabricius, W. V., & Khalil, S. L. (2003). False beliefs or false positives? Limits on children's understanding of mental representation. *Journal of Cognition and Development, 4*, 239-262.

Fivush, R. (2011). The development of autobiographical memory. *Annual Review of Psychology, 62,* 559-582.

Flavell, J. H. (1974). The development of inferences about others. In T. Mischel, Ed., *Understanding other persons.* Oxford: Blackwell, pp. 66-116.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance-reality distinction. *Cognitive Psychology, 15*, 95-120.

Flavell, J. H., Green, F. L. & Flavell, E. R. (1993). Children's understanding of the stream of consciousness. *Child Development, 64,* 387-398.

Flavell, J. H., Green, F. L. & Flavell, E. R. (1995). Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development, 60*, pp. 1-96.

Flavell, J. H., Green, F. L. & Flavell, E. R. (2000). Development of children's awareness of their own thoughts. *Journal of Cognition and Development, 1,* 97-112.

Flavell, J. H., Green, F. L., Flavell, E. R. & Grossman, J. B. (1997). The development of children's knowledge about inner speech. *Child Development, 68,* 39-47.

Flavell, J. H., Green, F. L., Flavell, E. R. & Lin, N. T. (1999). Development of children's knowledge about unconsciousness. *Child Development, 70,* 396-412.

Fodor, J. A. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind.* Cambridge, MA: The MIT Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong.* Oxford, England: Clarendon Press.

Fodor, J. A., Garrett, M., Walker, E., & Parkes, C. (1980). Against definitions. *Cognition, 8,* 263-367.

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

Fodor, J. A., & Lepore, E. (1992). *Holism: A shopper's guide.* Oxford, England: Blackwell.

Fricker, E. (2009). Is knowing a state of mind? The case against. In P. Greenbough & D. Pritchard (Eds.), *Williamson on knowledge*, pp. 31-60. New York: Oxford University Press.

Gallagher, S. & Meltzoff, A. N. (1996). The earliest sense of self and others: Merleau-Ponty and recent developmental studies. *Philosophical Psychology, 9,* 211-233.

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

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.

Gazzaniga, M. S. (1998). *The mind's past.* Berkeley: University of California Press.

Gennaro, R. J., Ed. (2004). *Higher-order theories of consciousness: An anthology.* John Benjamins.

Gennaro, R. J. (2012). *The consciousness paradox: Consciousness, concepts, and higher order thoughts.* Cambridge, MA: MIT Press.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis, 23,* 121-123.

Gopnik, A. (1988). Conceptual and semantic development as theory change. *Mind & Language, 3,* 197-217.

Gopnik, A. (2009). *The philosophical baby: What children's minds tell us about truth, love, and the meaning of life.* New York: Picador.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7,* 145-171.

Griffiths, P. (2009). The distinction between innate and acquired characteristics. In E. Zalta, ed., *Stanford Encyclopedia of Philosophy, Fall 2009 Edition.*

Harman, G. (1978) Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences 1*, 576–77.

Hedger, J. A., & Fabricius, W.V. (2011). True belief belies false belief: Recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development. *Review of Philosophy and Psychology, 2,* 429-447.

Horgan, T., & Woodward, J. (1985). Folk psychology is here to stay. *Philosophical Review, 94,* 197-225.

Kahneman, D. (2011). Thinking, Fast and Slow. London: Penguin Books.

Keysar, B., Lin, S. & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25-41.

Kriegel, U. (2003). Consciousness as intransitive self-consciousness: Two views and an argument. *Canadian Journal of Philosophy, 33,* pp. 103-132.

Kriegel, U. (2009). *Subjective consciousness: A self-representational theory.* Oxford: Oxford University Press.

Kurzban, R. & Aktipis, C. A. (2007). Modularity and the social mind: Are psychologists too self-ish? *Personality and Social Psychology Review, 11*, 131-149.

Levine, J. (2001). *Purple haze: The puzzle of consciousness.* Oxford: Oxford University Press.

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy, 50,* 249-258.

Lin, S., Keysar, B., & Epley, N. (2010). Refelexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology, 46,* 551-556.

Magnus, P., & Cohen, J. (2003). Williamson on knowledge and psychological explanation. *Philosophical Studies, 116*, 37-52.

McCormack, T. & Hoerl, C. (2005). Children's reasoning about the causal significance of the temporal order of events. *Developmental Psychology, 41,* 54-63.

McCormack, T. & Hoerl, C. (2007). Young children's reasoning about the order of past events. *Journal of Experimental Child Psychology, 98,* 168-183.

McGlynn, A. (2014). Knowledge first? Edinburgh, UK: Palgrave Macmillan.

Moll, H. & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology, 24,* 603-613.

Nagel, J. (2013). Knowledge as a mental state. In T. S. Gendler & J. Hawthorne, Eds., *Oxford studies in epistemology volume 4*, pp. 274-310. Oxford, UK: Oxford University Press.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83,* 435-450.

Nelson, K. & Fivush, R. (2004). The emergence of autobiographical memory: A social cultural developmental theory. *Psychological Review, 111,* 486-511.

Ornstein, P. A., Haden, C. A. & Elischberger, H. B. (2006). Children's memory development: Remembering the past and preparing for the future. In Bialystok, E. & Craik, F. I. M. (Eds.), *Lifespan cognition: Mechanisms of change.* New York: Oxford University Press.

Perner, J. (1988). Developing semantics for theories of mind: From propositional attitudes to mental representation. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 141-172). New York: Cambridge University Press.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: The MIT Press.

Povinelli, D. J. (2001). The self: Elevated in consciousness and extended in time. In Moore & Lemmon 2001, 75-96.

Povinelli, D. J., Landry, A. M., Theall, L. A., Clark, B. R. & Castille, C. M. (1999). Development of young children's understanding that the recent past is causally bound to the present. *Developmental Psychology, 35,* 1426-1439.

Putnam, H. (1962). The analytic and the synthetic. In H. Putnam, *Mind, language and reality: Philosophical papers, Volume 2*, pp. 33-69. Cambridge, UK: Cambridge University Press, 1975.

Ramsey, W., Stich, S. P., & Garon, J. (1990). Connectionism, eliminativism, and the future of folk psychology. *Philosophical Perspectives, 4,* 499-533.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*, 12-21.

Rose, D. (2015). Belief is prior to knowledge. *Episteme, 12,* 385-399.

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies, 49,* 329-359.

Rosenthal, D. M. (1993). Thinking that one thinks. In M. Davies and G. Humphreys (Eds.), *Consciousness*, pp. 197-223. Oxford: Blackwell.

Ruffman, T., Garnham, W., Import, A., & Connelly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology, 80*, 201-224.

Sacks, O. (1985). *The man who mistook his wife for a hat and other clinical tales.* New York: Simon and Schuster.

Sartre, J. (1943). Being and nothingness. Trans. H. E. Barnes. New York: Gallimard.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55,* 87-124.

Searle, J. (1992). *The rediscovery of the mind.* Cambridge, MA: MIT Press.

Searle, J. (2005). The self as a problem in philosophy and neurobiology. In T. E. Feinberg & J. P. Keenan (Eds.), *The lost self: Pathologies in the brain and identity*, pp. 7-19. New York: Oxford University Press.

Southgate, V. (2014). Early manifestations of mind reading. In S. Baron-Cohen, H. Tager-Flusberg & M. Lombardo (Eds.), *Understanding Other Minds*, 3rd Edition. Oxford University Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1-12.

Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind.* Cambridge, MA: MIT Press.

Vallar, G. & Ronchi, R. (2009). Somatoparaphrenia: a body delusion. A review of the neuropsychological literature. *Experimental Brain Research*, *192*, 533-551.

Van Gulick, R. (2006). Mirror mirror—Is that all? In Kriegel & Williford (2006), pp. 11-39.

Van Gulick, R. (2011). Consciousness. *Stanford Encyclopedia of Philosophy, Summer 2011 Edition.* Page numbers refer to the pdf version of the article.

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.

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35*, 245-275.

Wider, K. (2006). Emotion and self-consciousness. In Kriegel & Williford 2006, 63-87.

Williamson, T. (1995). Is knowing a state of mind? *Mind, 104*, 533-565.

Williamson, T. (2000). Knowledge and its limits. Oxford, UK: Oxford University Press.

Wittgenstein, L. (1953). *Philosophical investigations*. G. E. M. Anscombe and R. Rhees (Eds.). Oxford, UK: Blackwell.

Wittgenstein, L. (1969). On Certainty. Trans. D. Paul and G. E. M. Anscombe. Oxford: Basil Blackwell.

CURRICULUM VITAE

Joseph A. Hedger

September 2018

Contact Information:
address: Philosophy Department, Syracuse University
541 Hall of Languages, Syracuse, NY 13244-1170
e-mail: jhedger@syr.edu

Areas of Specialization:
Philosophy of Psychology, Cognitive Science, Philosophy of Mind, Philosophy of Language, Cognitive Development (Child Psychology).

Areas of Competence:
History of Analytic Philosophy (including Wittgenstein), Descartes, Philosophy of Biology, Developmental Psychology, Evolutionary Psychology, Applied Ethics, Ethics, Metaethics, Logic.

Education:
Ph.D., Philosophy, Syracuse University, defended 8/13/2018, expected Dec 2018.
Certificate of University Teaching, Syracuse University, May 2014.
M.A., Philosophy, Arizona State University, June 2010.
B.A., Philosophy, summa cum laude, Arizona State University, Dec 2007.

Academic positions:
Adjunct Professor, Philosophy, Syracuse University, 2015-
Adjunct Professor, Philosophy, Bryant & Stratton College, 2011-2018.
Teaching Assistant, Philosophy, Syracuse University, 2010- 2015.
Teaching Assistant, Philosophy, Arizona State, 2008- 2010.

Publications
2016
Hedger, J. A. (2016). Perceptual access reasoning: Developmental stage or System 1 heuristic? *Phenomenology and the Cognitive Sciences, 15,* 207-226.
Hedger, J. A. (2016). “Why Sainsbury and Tye Fail to Solve the Frege Puzzle.” *Linguistic and Philosophical Investigations, 15,* pp. 32-41.
2013
Hedger, J. A. (2013). “Meaning and Racial Slurs: Derogatory Epithets and the Semantics/Pragmatics Interface,” *Language and Communication, 33,* pp. 205-13.
Hedger, J. A. (2013). “New York is Just New York: An Account of Genuine Proper Names in Fiction.” *Review of Contemporary Philosophy*, *12*, pp. 43-59.

2012

Hedger, J. A. (2012). "The Semantics of Racial Slurs: Using Kaplan's Framework to Provide a Theory of the Meaning of Derogatory Epithets." *Linguistic and Philosophical Investigations, 11,* pp.74-84.

Hedger, J. A. (2012). "Modality, Theoretical Reductions, and the Lycan Objection: A Reply to Cameron." *Analysis and Metaphysics, 11,* pp. 75-89.

2011

Hedger, J. A., & Fabricius, W. V. (2011). "True Belief Belies False Belief: Recent Findings of Competence in Infants and Limitations in 5-Year-Olds, and Implications for Theory of Mind Development." *The Review of Philosophy and Psychology, 2(3),* pp. 429-447.

Presentations (APA):

"Frege's Puzzle Left Unsolved: A Reply to Sainsbury and Tye," American Philosophical Association, Pacific Division Meeting, San Francisco, 27 March 2013.

"The Truth about Racial Slurs: Using Kaplan's Framework to Provide a Semantic Theory of Derogatory Epithets," American Philosophical Association, Central Division Meeting, Chicago, 16 February 2012.

"New York is Just New York: An Account of Genuine Proper Names in Fiction," 84th Annual Meeting of the American Philosophical Association, Pacific Division, San Francisco, 1 April 2010.

Presentations (Professional, Invited):

"Slurs Express Contempt," Workshop on Pejorative Thought and Talk, University of Edinburgh, 9 December 2013.

"Why Sainsbury and Tye Fail to Solve the Frege Puzzle," Philosophy Forum Colloquium Series, SUNY Potsdam, 8 November, 2012.

"Do Infants Understand Belief?" Midwest Empirical and Theoretical Association for Philosophical Research Meeting, University of Illinois, Urbana-Champaign, 26 April 2012.

"The Truth about Racial Slurs: Using Kaplan's Framework to Provide a Semantic Theory of Derogatory Epithets," Philosophy Forum Colloquium Series, SUNY Potsdam, 17 March 2011.

"How We Understand Our Own Mental States: The Private Language Argument is Supported by Findings in Developmental Psychology," Philosophy and Religious Studies Society Spring Lecture Series, Arizona State University West, 11 March 2010.

"The Truth about Racial Slurs: Using Kaplan's Framework to Provide a Semantic Theory of Derogatory Epithets," Philosophy Colloquium Series, Arizona State University, 6 November 2009.

Presentations (Professional, Peer-Reviewed):

"Perception and knowledge in you and me: Young children's understanding of mental states in self and others" (poster presentation; C. Gonzales, W. Fabricius, A. Kupfer, S. Huerta, & J. Hedger) Society for Research in Child Development Biennial Meeting, Seattle, April 2013.

"How We Understand Our Own Mental States: The Private Language Argument is Supported by Findings in Developmental Psychology," (poster presentation) 36th Annual Meeting of the Society for Philosophy and Psychology, Portland, 9-10 June 2010.

"Cognition and Evolution: Models of the Mind in Brooks' Robot Architectures and Evolutionary Psychology," Cognitio 2009, Institut des sciences cognitives, Université du Québec à Montréal, 5 June 2009.

Courses as Sole Instructor:

PHI 317 Philosophy of Mind (Syracuse University, Fall 2014).
PHI 378 Minds and Machines (Syracuse University, Spring 2017, Spring 2018).
PHI 197 Human Nature (Syracuse University, Fall 2013, Spring 2014, Spring 2015, Spring 2016, Fall 2016, Spring 2017).
PHI 171 Critical Thinking (Syracuse University, online: Summer 2015, Fall 2015, Summer 2016, Fall 2016, Summer 2017, Fall 2017, seated: Summer 2018).
PHI 107 Theories of Knowledge and Reality (Syracuse University, Fall 2017, Spring 2018, Fall 2018).
PHIL 250 Practices in Analytical Thinking (Bryant & Stratton College, Summer 2011—Spring 2018; over 40 sections).
PHIL 317 Topics in Ethics, Philosophy & Religion (Bryant & Stratton College, Fall 2015).
PHIL 310 Logic and Reasoning (Bryant & Stratton College, Summer 2013, Fall 2014, Fall 2016, Summer 2017, Spring 2018, Summer 2018).
PHIL 222 Ethics of Health Care (Bryant & Stratton College, Summer 2016, Spring 2017, Fall 2017).
PSYC 101 Principles of Psychology (Bryant & Stratton College, Spring 2014).
SOSC 105 Career Development: Theory & Practice (Bryant & Stratton College, Spring 2015—Fall 2015).

Courses as Teaching Assistant:

PHI 101 Introduction to Philosophy (ASU Fall 2008, Fall 2009).
PHI 306 Applied Ethics (ASU Spring 2009).
PHI 314 Philosophy of Science (ASU Spring 2010).
PHI 107 Theories of Knowledge and Reality (Syracuse Fall 2011).
PHI 293 Ethics and the Media Professions (Syracuse Spring 2012).

Theory of Mind research:

Developmental Psychology lab of William V. Fabricius (Arizona State University), 2009-

Ph.D. Dissertation (Syracuse University)

"Seeing and Believing: Philosophical Issues in Theory of Mind Development"

Ph.D. Dissertation Committee

Robert Van Gulick (chair)
William V. Fabricius (Psychology)
Bernard W. Kobes
Kevan Edwards

Activities and associations:
American Philosophical Association: Student Member, 2008-
Society for Philosophy and Psychology: Student Member, 2009-
External Speaker Committee (Syracuse University): 2010-2011.
Internal Speaker Committee (Syracuse University): 2011-2012.
Graduate Students in Philosophy (ASU): President, 2008-2009; Vice Pres, 2009-2010.

Service: Reviews for the following journals:
Synthese
International Journal of Behavioral Development
Journal of Pragmatics
Language Sciences

Graduate coursework: (*= independent study, [a] = audit)
Philosophy of Mind and Cognitive Science:
Philosophy of Cognitive Science (Bernard W. Kobes), Fall 2008.
Evolution and Modularity (Bernard W. Kobes), Spring 2009.
Goldman's *Simulating Minds*[a] (Cheshire Calhoun), Fall 2009.
Recent Defenses of Dualism (Bernard W. Kobes), Spring 2010.
Fodor: Concepts, Reduction, and LOT* (Kevan Edwards), Spring 2011.
Theories of Concepts (Kevan Edwards), Fall 2011.
Consciousness and Cognitive Development* (Robert Van Gulick), Fall 2011.
Consciousness, Functionalism, and Reduction[a] (Robert Van Gulick), Spring 2012.
Theories of Consciousness[a] (Robert Van Gulick), Spring 2013.
Recent Work on Consciousness[a] (Robert Van Gulick), Fall 2018.
Psychology:
Cognitive Development of Theory of Mind (William V. Fabricius), Fall 2009.
Evolutionary Psychology (Douglas Kenrick), Fall 2009.
Developmental Psychology[a] (William V. Fabricius), Spring 2010.
Is Theory of Mind Modular?* (William V. Fabricius), Spring 2010.
Developmental Psychopathology[a] (Natalie Russo), Spring 2012.
Philosophy of Language and Logic:
Philosophy of Fiction (Peter A. French), Fall 2008.
Is Truth Relative? (N. Ángel Pinillos), Spring 2009.
Presupposition Failure and Truth Value Intuitions* (N. Ángel Pinillos), Fall 2009.
Logic and Language (Mark Brown), Fall 2010.
Reference and Representation (Kevan Edwards), Fall 2010.
Predication: Plurals and Mass[a] (Thomas J. McKay), Spring 2011.
History of Philosophy:
Wittgenstein's *Philosophical Investigations* (Peter A. French), Fall 2009.
History of Philosophy Proseminar (Kara Richardson), Fall 2010.
Frege: Philosophy of Language* (Andre Gallois), Spring 2011.
Metaphysics and Epistemology:
The Value of Knowledge (Steven Reynolds), Fall 2008.
Rational Disagreement (Stewart Cohen), Spring 2009.
Modality (Kris McDaniel), Fall 2010.

Metaphysics, Epistem., Mind & Language Proseminar (Andre Gallois), Fall 2011.
Ethics and Political Philosophy:
Metaethics and Moral Psychology (N. Ángel Pinillos), Spring 2010.
Moral and Political Philosophy Proseminar (Kenneth Baynes), Spring 2011.

Teaching workshops completed:
Creating and Using Rubrics for Grading, ASU Graduate College, 2008.
Teaching Large Classes: Keeping Students Engaged, ASU Center for Learning and Teaching Excellence, 2008.
Get Students Reading: Tips for Increasing Student Reading and Retention, ASU Center for Learning and Teaching Excellence, 2008.
I have also attended a number of Faculty Development seminars (Bryant & Stratton College, 2011-2015), and various workshops and presentations of the Future Professoriate Program (Syracuse University, 2012-2014).

M.A. Portfolio (Arizona State University)
1. "Preface: A Note on Methodology"
2. "New York is Just New York: An Account of Genuine Proper Names in Fiction"
3. "The Truth about Racial Slurs: Using Kaplan's Framework to Provide a Semantic Theory of Derogatory Epithets"
4. "The Delicate Balance between Ontology and Epistemology in the Mind-Body Problem: A Case for Modest Dualism as a Methodology in Psychology"

M.A. Committee
Bernard W. Kobes (chair)
N. Ángel Pinillos
Peter A. French

PROFESSIONAL REFERENCES

Robert Van Gulick
Professor, Philosophy
Syracuse University
rnvangul@syr.edu
315-443-5828

Bernard W. Kobes
Emeritus Professor, Philosophy
Arizona State University
kobes@asu.edu
480-965-5778

William V. Fabricius
Associate Professor, Psychology
Arizona State University
william.fabricius@asu.edu
480-965-9372

Kevan Edwards
Associate Professor, Philosophy
Syracuse University
Kedwar02@syr.edu
315-443-5821