5-2011

All Heads Gently Nodding: How Naturalism Dissolves the Problem of Other Minds

Devin Sanchez Curry
College of William and Mary

All Heads Gently Nodding
How Naturalism Dissolves the Problem of Other Minds

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Arts in Philosophy from
The College of William and Mary

by

Devin Sanchez Curry

Accepted for ______________________________________
(Honors, **High Honors**, Highest Honors)

____
Paul Davies, Director

____
Alan Goldman

____
Matthew Haug

____
Chris Ball

Williamsburg, VA
May 2, 2011

Table of Contents

Part I

The Problem of Other Minds

The epistemological problem of other minds – the hallowed philosophical quandary of justifying belief in the existence of other minds like one's own – is a problem about how humans come to know minds. Historically, most philosophers have broached the problem of other minds (henceforth the PoOM) entirely from their armchairs, without incorporating knowledge from relevant scientific theories. This thesis lays out what the science shows about how humans understand their own minds and other minds, and discusses what the evidence does to vindicate or deny leading philosophical accounts of attributions of mindedness. This evidence, drawn from work in psychology, cognitive science, neuroscience, and philosophy, strongly suggests that the PoOM is misguided and confused from the get-go and that there is no greater problem with belief in other minds than belief in one's own.

More specifically, my project in this essay will be to augment a fairly set-in-stone philosophical route to dissolving the PoOM with scientific evidence. When argued for from a conceptual analytic stance (without reference to findings in science) this route is weak; but when argued for from a progressive naturalistic stance it is much more plausible. The conceptual analytic approach, as exemplified by P.F. Strawson and his fellow post-Wittgensteinian philosophers, will be detailed in Chapter 2. My progressive naturalistic approach will be detailed in Chapter 3. I will argue that the sort of attack on the PoOM the post-Wittgensteinians endorse is difficult to defend without the help of scientific evidence, but can be robustly defended with science on its side. But first, in Chapter 1, I motivate this project by briefly reconsidering certain key features of the Cartesian account of mindedness.

Chapter 1

Reconsidering the Cartesian Viewpoint

Rene Descartes held that he could be absolutely certain of his own existence before acquiring knowledge about anything else. The *cogito*, his infamous declaration that “I think therefore I am”,[1] takes knowledge of the existence of one's own mental life to be indubitable. One comes to know oneself as minded not through *a posteriori* exploration, but through an automatic act of pure reason. Erstwhile, dualism, the lynchpin of the Cartesian account of mind, introduces a radical divide between mind and body. This split is not trivial; it is powerful and basic. It is not just that minds can be distinct from bodies. Rather, they are composed of a different kind of substance altogether. The split between mind and body is also, for Descartes, a given. It is not contingent on anything, but instead an objective logical divide. It is logically impossible for a mental thing and a physical thing to be one and the same, just based on what the concepts “mind” and “body” entail. That is, part of the definition of mind is 'not bodied' and part of the definition of body is 'not minded'. Dualism thus rigidly distances the (mental) subject from the (bodily) world.

Notice the implications of Descartes' metaphysical dualism for his account of

[1] Descartes (1644) Part 1, Article 7

epistemology. Knowledge, for Descartes, consists in the closing of the distance caused by substance dualism. Knowledge is the subject coming to understand the world. But, even if the subject becomes very knowledgeable about her world, even if that distance becomes very close, the dualist must maintain that the two remain isolated. It is always feasible, even when the subject comes to know and interact with the world, to distinguish subject and world without any overlap.

Accepting the logical nature of the mind/body split makes the PoOM inescapable, and perhaps unconquerable. The PoOM is the problem of attaining knowledge of other subjects. But, given this Cartesian framework, such knowledge is at best inferable, and maybe impossible. For the only thing a subject can know as a subject is herself. We come to think we know other people as minded by interacting with them in the world, not by pure reason. Given a radical logical divide, we thus cannot truly come to know other people as minded. Acquiring knowledge involves getting information from the world and this information must always be physical in nature. And thus, if we submit to the Cartesian framework, we cannot acquire knowledge of other minds.

The assumption of a general Cartesian framework, including some sort of logical divide between mind and body, is made not only by dualists, but by the vast majority of philosophers who have ever worked on the PoOM. From Locke and Malebranche in the late 17th and early 18th centuries to Searle and Galen Strawson in the late 20th and early 21st centuries, the history of the PoOM has, in large part, been the history of reconciling the Cartesian viewpoint with other minds.[2] Descartes may have had good reason to push his viewpoint, given the relative dearth of extant scientific evidence regarding

[2] See Avramides (2001)

mindedness available in the early 17th century. Most of his successors, however, and especially those philosophers working in his tradition today, have had no such excuse. It is my methodological commitment that, given the enormous progress in human understanding of the world as a direct result of the advent and application of the scientific method, it is a mistake to ignore that understanding. Rather, by incorporating our best scientific knowledge into our philosophical reflections, we stand to enrich and expand the breadth of our knowledge and the precision of our inquiry.

Moreover, I believe that there is good reason to assume that Descartes himself, were he alive, well, and brought up to speed in 2011, would renounce his own philosophy of mind. Descartes was, perhaps above all else, a good scientist. Despite being bogged down by theological commitments, strong naturalistic tendencies shine through much of Descartes' writings. For example, in “Descartes' Philosophy of Science”, Desmond Clarke notes that “for Descartes, to explain a natural phenomenon is … to construct a mechanical model of how the phenomenon in question is caused.”[3] In Chapters 6 and 7 I describe in detail a mechanical model that, I propose, causes the phenomenon of humans understanding themselves and others as minded. This model explicitly rejects Descartes' core views. Nonetheless, I cannot help but think it plausible that, were Descartes to examine the relevant evidence, he would be swayed by my argumentation and seriously consider giving up his convictions.

Thus this thesis will not assume that a radical Cartesian divide between mind and body exists. I will also not assume the existence of direct and transparent access to one's own mental states. There is much better reason to believe that the evidence will support

[3] Clarke 282

introspective access to one's own mind than that it will support dualism, but even so, for reasons more fully explicated in Chapters 3 and 5, it is important that I withhold judgment until that evidence comes to light. For now, suffice to say that an integral part of reconsidering the Cartesian framework around the PoOM is rejecting the *a priori* assumption of mental transparency.

I hope that my unwillingness to assume dubious concepts prior to analysis comes off as somewhat trivial. After all, the job description of a philosopher importantly includes examining basic concepts. Furthermore, my challenge that philosophers are prone to assuming dubious concepts is hardly a new one. Francis Bacon's *New Organon* (1620) is a clear precursor to the progressive naturalistic methodology I endorse in Chapter 3 and employ throughout this thesis. In the *New Organon,* Bacon writes that “the idols and false notions which are now in possession of the human understanding, and have taken deep root in there, not only so beset men's minds that truth can hardly find entrance, but even after entrance is obtained, they will again, in the very insaturation of the sciences, meet and trouble us, unless men being forewarned of the danger fortify themselves as far as may be against their assaults.”[4]

I will spend a good portion of this thesis explaining how we might fortify ourselves against the assaults of dubious concepts like mental transparency because I take many philosophers, both contemporary and historical, to have conveniently neglected their jobs in order to remain under the thrall of Bacon's idols. However, perhaps unlike Bacon, I will be fully willing to reclaim previously dubious concepts if those concepts are vindicated by our best sciences.

[4] Bacon XXXVIII

Chapter 2

Two Problems of Other Minds: The Post-Wittgensteinian Approach

There are two problems of other minds discussed in the literature today. The better known version is the epistemological problem: the problem of justifying belief in the existence of other minds like one's own. This worry arises whenever a person believes that other minds exist, much like she believes her own mind exists, but runs into a difficulty justifying that belief that she does not run into when justifying belief in her own mind. In detail, the argument that sets up the classic epistemological PoOM runs as follows.

P1 One believes that oneself is minded.

P2 One believes that others are minded.

P3 The concept of mindedness at work in P1 and P2 is univocal.

P4 One justifies belief in one's own mind by virtue of direct access to that mind.

P5 One lacks such direct access to other minds.

C One must find another way to justify belief in other minds.

The above is a valid argument, and its conclusion is the epistemological PoOM. So, if the premises P1-P5 are all true, then the PoOM is a real epistemological worry. In particular, the conjunction of P4 and P5, in light of P3, casts doubt on the belief asserted by P2. The way in which one justifies mindedness is direct access thereto, but one only has direct access in the 1st person case. Thus, unless one can come up with an alternate satisfactory justification, one's belief in other minds is dubious.

My rejection of the Cartesian viewpoint in Chapter 1 is far from new to the ongoing debate about the PoOM. Indeed, freed from the rigidity of Cartesian categories, contemporary analytic philosophers often argue that the epistemological PoOM simply does not occur. To make this point clear, they sometimes posit and work through a different version of the PoOM. This second problem is the conceptual PoOM, and it is the problem that some contemporary philosophers take to be the real issue behind the classic epistemological PoOM.[5] The conceptual PoOM is the problem of detailing the concepts that are at work in the epistemological PoOM. Proponents of the conceptual problem note that the epistemological problem is confounding precisely because, as noted in Chapter 1, indoctrinated Cartesian concepts make it impossible to solve. So, because wrangling with the PoOM is futile, they attempt to sidestep wrangling and illustrate why the PoOM is itself futile.

The most widespread versions of the conceptual PoOM deal directly with P3 in the argument laid out above. They question whether the concept of mindedness really is univocal from one case to the next. One such version is the problem of generality. The problem of generality holds that in order to question the mind of another, one must have a concept of mind general enough to apply to others as well as oneself. Given the varying evidence at work in 1st person and 3rd person ascriptions of mindedness, and the plausibility that those ascriptions differ in methodology, it is dubious whether such a general concept exists. Another, related, version of the conceptual problem is the problem

[5] Avramides 219

of unity. The problem of unity suggests that, for the PoOM to occur, the concept at play in ascribing mental states to others must be the very same as the concept at play in self-ascription.

The problems of generality and unity both straightforwardly originate in Wittgenstein's (1953) discussion of private language in the *Philosophical Investigations*.[6] Wittgenstein's prime example is pain. The problem of generality questions whether my concept of pain is general enough to refer both to my own sensation of being whacked upside the head and to observing you react to being whacked upside the head. In my case, there is a distinct qualitative pain. In your case, from my viewpoint, there is only observation of behavior indicating such a feeling. Is my concept of pain general enough to cover my raw qualitative pain as well as whatever is indicated by your startled behavior? The problem of unity, meanwhile, questions whether when I actually state "I am in pain" and "you are in pain" I am referring to the same concept. When I claim "I am in pain" I know that I am referring to a real present ache. When I say "you are in pain", however, I can only refer to your behavior which indicates a supposed ache. Given my degree of certainty, or lack thereof as the case may be, can the concept "pain" really be univocal as it works in these two statements?

As just noted, the epistemological PoOM, per P3, premises the generality and unity of the concept of mindedness in the 1st and 3rd person cases. It also, per P4 and P5,

[6] Two notes. First, while Wittgenstein's pain example originates and nicely illustrates the problems of generality and unity, Wittgenstein is not himself troubled by the problems. He explains why in the private language argument in the *Philosophical Investigations*. Second, the post-Wittgensteinian analysis of the conceptual PoOM (as explicated in this chapter) differs substantially from Wittgenstein's own analysis in the *Philosophical Investigations*. Those philosophers whom I term post-Wittgensteinian in this thesis are post-Wittgensteinian in that they follow in his footsteps and are inspired by his work, but they do not always argue for the same conclusions.

premises a lack of unity in access to this concept of mindedness in the 1st and 3rd person cases. To combine these two facts is to say that the PoOM takes mindedness to be such that I and others can have it, but maintains that the way in which I come to understand myself to have it differs importantly from the way in which I come to understand others to have it.

While the conceptual PoOM can be phrased in many ways – as the problem of generality, the problem of unity, et cetera – it essentially boils down to a challenge to define the concepts that are assumed in the epistemological PoOM and then to determine if those embedded concepts make the epistemological problem internally consistent. That is, to determine whether or not the PoOM contains contradictory or false premises. Conclusions obviously vary, but one type of answer consistently arises from philosophers with a Wittgensteinian bent. This post-Wittgensteinian response is unique in that it addresses the conceptual PoOM not by denying P3, but by questioning P5. P.F. Strawson's account of the conceptual problem in "Persons" is perhaps the most prototypical of this type of answer.

In "Persons", P.F. Strawson argues that by examining usage in language and rejecting Cartesian intuitions, the conceptual PoOM can be solved, as the concept of mind is found to be univocal in the 1st and 3rd person cases. However, Strawson takes the process of vindicating P3 to coincidentally undermine P5. He takes humans to have the same sort of direct access (but not Cartesian access) to other minds as to one's own. Thus, in solving the conceptual problem, he dissolves the epistemological problem. Before getting into the details of how Strawson's argument runs, though, it is useful to take a step back and make clear his general methodology.

“Persons” is a chapter of Strawson's book *Individuals*. In *Individuals*, Strawson engages in his unique version of Oxford-style ordinary language philosophy: a practice he terms 'descriptive metaphysics'. Descriptive metaphysics (unlike most metaphysics, which Strawson terms 'revisionary'), is the endeavor of describing “the actual structure of our thought about the world.”[7] While individual languages and terminologies can vary throughout time and place, Strawson maintains that there exists a “massive central core of … categories and concepts which, in their most fundamental character, change not at all.” [8] There is a fundamental human grammar through which fundamental human concepts are expressed, and descriptive metaphysics involves determining the precise role of these concepts in this grammar. It is Strawson's methodological commitment that “the reliance upon a close examination of the actual use of words is the best, and indeed the only sure, way in philosophy.”[9] He thus examines the actual human use of words such as 'mind', 'body', 'person', 'action', and 'I' in order to attack the PoOM.

Strawson does not take a Cartesian logical divide between mind and body to exist. Indeed, for Strawson, the opposite is true. By virtue of human grammar, it is logically impossible to divide the concepts of 'mind' and 'body'. One's personhood, which is more basic (Strawson uses the term 'primitive') than either mind or body, consists of both. A person is a thing that has mental states and physical states, not distinctly, but in conjunction as states of that person. Particular states (mental or physical) only exist insofar as they are states of a particular person. The logical impossibility, then, is for a

[7] Strawson 9
[8] *ibid* 10
[9] *ibid* 9

particular state to belong to anyone but the person to whom the state does in fact belong. We attribute 'has a mind' and 'has a body' to persons, but we never attribute 'is a person' to a mind or a mindless body. There are not, Strawson maintains, “two uses of 'I', in one of which it denotes something which it does not denote in the other.”[10] It is therefore not that the person encompasses the physical 'I' and the mental 'I', but that the person is the only 'I', comprising physicality and mentality. This unified 'person' is in fact the only way we conceptualize people through ordinary language. For Strawson, this fact is overwhelming clear simply on the basis of observable patterns of linguistic usage.

Strawson takes it to follow from this assertion that the primitiveness of 'is a person' holds in all cases. It holds when I attribute mindedness to myself just as when I attribute mindedness to others. If others are Cartesian subjects, logically divided from the world in the same way as oneself, then one can confidently ascribe a mind always to oneself but never to others. But, if others are complete unified persons like one is a person, if, as Strawson holds, mental states are only ascribed to a person insofar as physical states are ascribed to that same person (for this dual-ascription is what it is to be a person), then the concept 'person' is logically prior to the concept 'mindedness'. Further, Strawson takes the concept 'person' to only be useful insofar as there exist other persons. It is actually a necessary condition of the ability to take oneself to be minded that one can take others to be likewise minded. Otherwise, it would be a fruitless ascription. There is no point in ascribing a certain mindedness to myself in particular if I do not take other beings to whom I might ascribe it to exist.

But it is more than useless. The very structure of the concept of mindedness, as

[10] *ibid* 98

evidenced by its use in common language, makes it so that it must be used in both 1st and 3rd person ascriptions. Strawson writes that “there is no sense in the idea of ascribing states of consciousness to oneself, or at all, unless the ascriber already knows how to ascribe at least some states of consciousness to others. So he cannot argue in general 'from his own case' to conclusions about how to do this; for unless he already knows how to do this, he has no conception of *his own case*, or any *case*, i.e. any subject of experiences.”[11] [12] The concept is fundamentally univocal. One cannot warrant ascribing mindedness to anything without understanding it as something that both oneself and others have.

How, then, do we know to whom we should ascribe minds? Strawson argues that we ascribe personhood by recognizing action. There is a certain kind of bodily movement, called action, that is indicative of agency. Action is the intentional minded behavior of a subject. Persons act. And persons ascribe personhood to others by observing them acting similarly. Likewise, persons observe other persons acting, and ascribe personhood to themselves by acting similarly. 'Personhood' is thus a concept not arrived at in the 1st person case and then attributed to others, nor a concept arrived at in the 3rd person case and then attributed to oneself. Rather, we come to know ourselves and others as persons simultaneously through interaction. Humans necessarily see themselves and others as persons via interaction, and see themselves and others as minded via personhood.

So, for Strawson, as the PoOM assumes, the concept of mind is general and

[11] Strawson 106

[12] Strawson's italics.

univocal. But, counter the PoOM, the way in which one comes to know one's own mind and other minds is also univocal. So, if one doubts the existence of other minds, one must equally doubt the existence of one's own mind. The PoOM becomes the problem of any minds. Moreover, as persons employing the grammar we use in everyday life, it is impossible to actually doubt the existence of minds.

Anita Avramides, in her 2001 book *Other Minds*, explains how Strawson's account meshes with Donald Davidson's account[13], and implicitly takes her hybrid-account to be the ultimate post-Wittgensteinian answer. For Avramides, action, or 'behavior proper' as she terms it, is that which keeps there from being a divide between subject and world. Humans understand behavior proper, both their own and others', by virtue of occupying what she refers to as the 'lived position'.[14] According to Avramides, the PoOM occurs to philosophers because, in accepting Cartesian categories, they distance themselves from real everyday life. Wrongheaded philosophers assume the stance of Descartes, in his armchair, alone in a room for days on end. From this philosophical starting point, it is easy to adopt the belief that the subject is merely a mindedness. After all, one has hardly any need to believe in other minds when there are currently no other minds that need addressing.

But Avramides wholeheartedly rejects this belief and starting point. She argues that in real life – the life of the non-philosopher – the subject is a whole person: body and mind. The lived position is a philosophical starting point that takes the foundation of inquiry not to be an isolated rationality, but rather a person interacting with other persons

[13] See Davidson (1997)

[14] Avramides 229

in everyday life. From this position, Avramides holds, it is plainly and naturally evident that persons are united by behavior proper.

If the post-Wittgensteinians (as exemplified by Strawson and Avramides) are correct, then the conceptual PoOM is solved, and P3 of the epistemological problem is vindicated. The concept of mindedness at work in P1 and P2 is in fact univocal. Nonetheless, the epistemological PoOM does not occur, as another of its premises, namely P5, is false. Other minds are not a special case; one, through interaction, has the same sort of direct access to the mindedness of others as to her own. I am not going to address objections to the post-Wittgensteinian conclusion now because, while I take myself to have adequately lain out the evidence and arguments that the post-Wittgensteinians give for their conclusion, I do not believe that this evidence and argumentation is sufficient to support the conclusion. Indeed, in Chapter 4 I will provide an example of how well-accepted science casts aspersions on the account.

I am, however, inspired by the post-Wittgensteinian approach. I believe that attempting to dissolve the epistemological problem by solving the conceptual problem can be a fruitful project. So, rather than defending the truth of the above conclusion, which is probably indefensible as it stands, I will attempt to bolster the project. I will end up asserting, counter the post-Wittgensteinians, that P5 is actually true, but that it is trivially true, as P4 is false.

The post-Wittgensteinians, at times, seem to crave reference to psychological literature. For instance, Strawson writes that "it is easier to understand how we can see each other, and ourselves, as persons, if we think first of the fact that we act … in accordance with a common human nature"[15] and Avramides notes that "understanding the kind of beings that we are and the way we act, with others, in the world can help to make the concept of mind that we operate with intelligible to us."[16] Nevertheless, neither Strawson nor Avramides use science, that endeavor which produces bountiful information about "common human nature" and "the kind of beings that we are", to support their conclusions.

Strawson takes himself, in *Individuals,* to be engaging in descriptive metaphysics. While descriptive metaphysics may be a worthy pursuit, I believe that if an accurate description of the world is one's aim, then science should be the first place one looks. If, as Kant claims, the name 'metaphysics' is not accidental, then descriptive metaphysics should come only after descriptive physics. Methodological scientific inquiry into nature should not necessarily be more highly valued than philosophical inquiry, but, as I will argue in Part II, it should be taken into account at the onset of philosophizing.

[15] Strawson 112

[16] Avramides 290

Part II

A Naturalistic Methodology

Part I introduced the conceptual and epistemological problems of other minds, and explained how the post-Wittgensteinians go about rejecting the Cartesian viewpoint. Namely, Strawson and Avramides appeal to social interaction and usage in language to solve the conceptual problem and dissolve the epistemological problem. While I do not take the post-Wittgensteinian dissolution to be adequate, I am inspired by the idea of dissolving the epistemological problem through solving the conceptual problem. Thus, Part II introduces my own methodology that I will eventually take to accomplish the same end. Chapter 3 explicates this methodology, which focuses on refusing to assume dubious concepts. Then, Chapter 4 examines the ways in which science can and cannot be fruitfully applied to skeptical philosophical debates.

Chapter 3

Embracing Naturalism

The problem of other minds is, at its core, a problem about minds. Minds, if they exist, and whether or not they are explicable in purely physical terms, are real natural things. One of the best (if not the only reliable) ways of studying natural things is the scientific method. Postmodernist worries aside, I am confident that at the very least all of the theorists with whom I engage in this thesis (including – perhaps especially –

Descartes[17]) would agree that science is a legitimate means by which progress in the understanding of minds might be achieved. Indeed, in psychology, cognitive science, and neuroscience we have sciences devoted entirely to the study of minds. Thus, as philosophical discussion about minds takes place, we should make clear what facts are already given to us by science.

Philosophers who interpret science – at least those who have a command of the relevant compelling evidence and theorize not by appropriating that evidence, but by analyzing it – are not doing amateur science. Rather, philosophers who do not interpret science are doing naive philosophy. This much I take to be uncontroversial among a philosophical community for which naturalism is increasingly the norm. However, such alleged naturalistic commitments are concurrently unpracticed. Most philosophers address the PoOM not only without reference to the relevant scientific evidence, but largely in ignorance thereof.

My methodological commitments are almost entirely indebted to Paul Sheldon Davies, and particularly to his book *Subjects of the World: Darwin's Rhetoric and the Study of Agency in Nature*. *Subjects of the World* is essentially a manifesto for what Davies terms 'conceptual progressivism', a naturalistic orientation towards philosophical inquiry that expects old concepts to be unsettled by cutting-edge science. For the conceptual progressivist, inquiry is exploration. Her goal is not to preserve concepts wantonly but to make sure all of her concepts, whether historically or psychologically entrenched or new, conform to the world (as studied scientifically).

For Davies, Darwin is the model progressive naturalist. Darwin's theory of

[17] See Descartes (1949 and 1964); especially his discussion of the pineal gland as the “seat of the soul.”

evolution by natural selection did not, like its contemporaries, attempt to uphold creationism when the science overwhelmingly showed that it was a dubious concept. Indeed, he not only refused to let any entrenched concepts compel him to interpret his scientific evidence in any certain way, but moreover expected his formerly entrenched concepts to be upset by the evidence. Darwin also, as any progressive naturalist must, understood that the new conceptual categories he was proposing might be jarring; for example, he realized that people had difficulty conceptualizing large changes being the product of millions of slight variations over millions of years.[18] He interpreted the scientific evidence in light of only itself, and from the world devised those concepts worth believing.

The progressivist's foil is the 'conceptual conservative'. Conceptual conservatives are, as the name suggests, simply thinkers who attempt to conserve hallowed concepts, whether those concepts are passed on by tradition, religion, history, science, or intuition. According to Davies, there are two types of conceptual conservatism. The first focuses on the preservation of "any concept that appears important within some well-developed scientific theory"[19], and the second focuses on the preservation of "any concept that appears important within our general worldview."[20] Both types focus, then, on the preservation of certain concepts as we pursue knowledge. They do so because of a couple of misguided heuristics that have long held sway in theology and philosophy. The first of these heuristics is that the longer the historical roots of a concept, the more likely it is a good concept. The second is that the more venerated a concept, the more important it is to

[18] Darwin 34

[19] Davies 24

[20] *ibid*

integrate with new knowledge. In other words, the conceptual conservative premises concepts based on neither *a priori* nor *a posteriori* argumentation, but instead merely because they are concepts which tend (and have tended) to be premised in religious or philosophical debates.

Particularly egregious conservatives are what Davies terms 'conceptual imperialists'. Conceptual imperialists not only try to save outdated concepts in light of contrary scientific evidence, but allow those concepts to dominate that evidence. They reject any evidence that is counter to their venerated concepts. For example, the philosopher Roderick Chisholm is an imperialist with regards to free will.[21] He attempts not only to preserve the concept of humans as unmoved movers in light of science, but indeed is willing to distort or disregard science entirely in order to uphold his concept.[22] This attempt to preserve venerable concepts is not, in and of itself, a bad thing. However, the attempt to preserve dubious concepts, no matter how venerable, is.

Theology and much of philosophy are conservative or imperialistic in that, unlike Darwin, they engage in the concept location project for dubious concepts. The concept location project involves taking preset concepts and figuring out how they fit with science. The problem is that these concepts that are assumed to exist are often dubious in one or both of two ways. They may be dubious by descent in that – like Francis Bacon's idols of the marketplace and theater[23] – they arose from a now defunct worldview. Or they may be dubious by psychological role in that – like Bacon's idols of the tribe and cave[24] – they arise from a human psychological module that is apt to generate false-

[21] Chisholm 25

[22] Davies 26

[23] Bacon XLIII and XLIV

[24] Bacon XLI and XLII

positives. Conservative and imperialistic theologians and philosophers attempt to preserve these dubious concepts whether or not they are supported by the evidence.

A concept is dubious by descent if it arose from a worldview no longer regarded to be true, especially if that worldview is theological or political in any way. This is not to claim that theological or political views are necessarily false, but only that their blind preservation has no place in scientific or philosophical inquiry. For, as Bacon notes, concepts dubious by descent “plainly force and overrule the understanding, and throw all into confusion, and lead men away into numberless empty controversies and idle fancies.”[25]

So, for example, Chisholm's concept of libertarian free will is dubious by descent. The concept has strong roots in Western theological and philosophical history, tracing back to Descartes (who claimed that men have an "infinite" will) and firmly to Judeo-Christian theology. There is nothing in the science which plausibly supports libertarian free will, and very few avowed naturalists buy into Cartesian or Judeo-Christian doctrine today. So, we should not, as Chisholm does, make it a condition of adequacy on our theorizing that we preserve this concept of free will. We should bracket the concept – not throw it out, but set it aside – and, as inquiry progresses, note how the evidence independently supports or rejects it. This is progressive in that it expects and allows the concept of free will to be unsettled by the evidence, does not allow the concept to be preserved if it does not conform to our understanding of the world, and yet still leaves room for the concept to be vindicated if that is what the evidence demands.

A concept is dubious by psychological role insofar as intuition is our strongest

[25] Bacon XLIII

available reason to uphold it. An important lesson from contemporary cognitive psychology is that some of the human mind's most basic conceptualizing capacities lead us to believe things that are false. For various evolutionary reasons, we have psychological modules underwriting consciousness that are motivated by factors other than the truth, and thus generate an abundance of false positives. In defending this point Davies obviously makes reference to a considerably deeper reservoir of scientific evidence than was available to Bacon in the late 16th and early 17th centuries. But Bacon notices the dubiousness of concepts propagated by psychological role all the same, writing that "it is a false assertion that the sense of man is the measure of things. On the contrary, all perceptions both of the sense and of the mind are according to the measure of the individual, and not according to the measure of the universe. And the human understanding is like a false mirror, which, receiving rays irregularly, distorts and discolors the nature of things by mingling its own nature with it."[26]

Take, for example, what novelist David Foster Wallace terms the 'default setting': my overriding intuition that I am the exact conceptual center of the universe.[27] All of my experiences are of things and events in relation to me, and only in relation to me. I thus naturally, by virtue of my psychological relationship to the world, have a concept of myself as the most important thing in the universe. This is, of course, false. And so too might be other (less obviously wrongheaded) concepts that play important roles in our psychology. Thus, we should not make it a condition of adequacy on our philosophical theorizing that we preserve such concepts, but instead should seek to understand the role

[26] Bacon XLI

[27] Wallace 38

of the concept and thereby figure out if it is worthwhile. This is progressive in that it expects our capacities to be deceptive, and seeks to figure out just what brings concepts about before making a judgment on whether or not those concepts are veridical.

Davies proposes directives for naturalistic inquiry which help philosophers avoid conservatism and pursue progressivism. Two of those directives, precisely those directives that order us to avoid dubious concepts not justified by science, run as follows.

> Concepts dubious by descent (DD): For any concept dubious by virtue of descent, do not make it a condition of adequacy on our philosophical theorizing that we preserve or otherwise "save" that concept; rather, bracket the concept with the expectation that it will be explained away or vindicated as inquiry progresses – as we analyze inwardly and synthesize laterally.[28]

> Concepts dubious by psychological role (DP): For any concept dubious by psychological role, do not make it a condition of adequacy on our philosophical theorizing that we preserve or otherwise "save" that concept; rather, require that we identify the conditions (if any) under which the concept is correctly applied and withhold antecedent authority from that concept under all other conditions.[29]

It is important to stress that if the science does bear out dubious concepts, then the progressive will gladly reclaim them. So if good scientific evidence emerges that Devin

[28] Davies 42

[29] Davies 44

Sanchez Curry is the most important thing in the universe, then the progressive should happily accept that fact. But until that empirical evidence emerges – until that concept is no longer extremely dubious – it should not be taken as true.

In Bacon and Descartes' day, thinkers had access to significantly less empirical evidence about the world than they do today. Even so, as glossed in Chapter 1, Descartes labored to understand the natural world on its own terms. The same, unfortunately, cannot be said for much of the grand tradition of thought which followed in his footsteps. The Cartesian viewpoint on mindedness, as it is held today, is dubious by both descent and psychological role.

That dualism is dubious by descent is strikingly clear. It is a concept that arose from various deeply theological worldviews that assume that humans possess immaterial (and immortal) souls. Whether contemporary dualists are following Plato, Averroes, Aquinas, Descartes, or their local preacher, they are attempting to preserve a concept which the science simply does not bear out. It is conceivable that a radical mind/body split exists, but it is highly dubious, both because there is nothing concrete speaking for it in the science, and because it is a concept that is often propagated in fulfillment of purely dogmatic motives. In Chapter 5 I will explain how another integral part of the Cartesian account of mindedness – the mental transparency assumption – is dubious by both descent and psychological role.

Dualism, insofar as it has any serious influence today, is a theory latched onto by conceptual conservatives and imperialists attempting to preserve and propagate deeply dubious faiths and intuitions. By explicitly bracketing the concept of substance dualism, the post-Wittgensteinians are taking an important step away from conservatism and towards progressivism. The goal of this thesis, then, could be construed as guiding philosophers working on the PoOM today the rest of the way into a progressive naturalist mindset.

Post-Wittgensteinian thought, while distancing itself from the Cartestian viewpoint, still attempts the philosophical PoOM without recourse to the scientific facts at its disposal. Post-Wittgensteinian philosophers thus run the risk of themselves pushing concepts that are dubious by descent and/or psychological role. Further, unlike Descartes, they do so in an era in which they have ample access to a vast and ever-expanding wealth of scientific literature. To ignore this literature is a grave mistake, and it is difficult to warrant taking seriously any philosopher – indeed any theorist in any field – who does so. Thus, so that we might be able to seriously consider their position, the remainder of this thesis intends to correct that mistake on the part of the post-Wittgensteinians. In Chapter 4 I will examine an application in which the science appears to cast aspersions on the post-Wittgensteinian account, but end up marking it as a prime example of the wrong way of going about applying science to philosophical debates. In Part III I will go about applying science to philosophy the right way: by framing the conceptual PoOM as a target of naturalist inquiry, and replacing the dubious concepts that cloud the topic with naturalistically acceptable concepts. Then, in Part IV, I will examine how well the post-Wittgensteinian conclusions map onto the naturalistic conclusions.

Chapter 4

Science Paves the Conceptual Path

The problem of other minds is not a problem if there is no belief in other minds to justify. But before I go on, I should distinguish between two types of belief. The first is what I term 'explicit belief'. We explicitly believe things when we have good reason to think they are true, based on consciously considered evidence. However, we also have 'implicit beliefs'. A belief is implicit when we believe it whether or not we have considered evidence that justifies believing it. Some implicit beliefs may be unavoidable, while others may be open to some kind of control.

Now, there is fairly good reason to assume I believe in other minds in one way or the other. After all, I interact with others everyday and each of those interactions turns on the coacknowledged (i.e. at least apparently implicitly believed) fact that each of us has a mind. That is to say that each of us has, and acknowledges that the other has, those mental states such as intentions, beliefs, desires, et cetera, that philosophers call propositional attitudes. Propositional attitudes are those mental states that indicate a person's relationship to a proposition. I certainly at least acknowledge that others have attitudes about things in the world. But such anecdotal evidence immediately raises a red flag. Does the acknowledgment of mental states in the above interaction entail belief in mental states, or might that acknowledgment be made in lieu of actual belief because acknowledgment of mental states is useful for the interaction? It seems straightforward that if one does not really believe in other minds, and instead only treats others as minded because it is a useful thing to do, then the PoOM does not occur. P2 of the argument laid out in Chapter 2 is false: there is no real belief in other minds to justify.

Prima facie this may seem like a queer concern. It is a question about what I believe, and it is frequently assumed, among philosophers and laypersons, that I am not only the eminent authority on what I believe, but am flawless in attributions of belief to myself. The assumption is that it is a truism that I believe what I think I believe. In Chapter 5 I will address why this assumption is dubious. For now it will suffice to explore the question of whether I believe in other minds because I believe them to truly exist (as opposed to treating other minds as existent on instrumental grounds) from a third person perspective, without appeal to my own intuitions.

I will argue that in light of psychological research on mindreading[30], the possibility that belief in other minds is driven by utility is not a good challenge to the PoOM. But the challenge will not be dismissed because the utility claim is dismissed. Rather, it is plausible that one implicitly believes in other minds for mainly instrumental purposes, but that this possibility only strengthens the PoOM.

Most psychologists today hold that humans cannot help but see other animated objects as minded. Specifically, they claim that the human mind comprises, in part, a mindreading component: a set of capacities that incline humans to see themselves and others as minded. In essence, mindreading constitutes the biological impetus for the human ability to attribute, and thus believe in, minds. How and on what this mindreading component works is, unlike its existence in some form or another, a contentious issue. I

[30] Following Peter Carruthers, I use the more recently in-vogue term "mindreading" rather than the more traditional "theory of mind" to make it clear that the term is useful no matter which stance one takes in the theory/simulation debate in psychology.

will fully explicate my stance on the mindreading debate in Chapter 6.

But for now, to return to the issue at hand, I pose the following question. Assuming the existence of an evolved mindreading system, do I believe that you are minded or do I merely treat you as minded because it is useful to do so? The answer is that this is a false dichotomy: I probably do both. I treat you as minded because it is useful to do so and, because it was similarly useful in the evolutionary history of our species, my mind forces me to implicitly believe you are minded. I really do honestly believe you are minded. Nevertheless, deeper down, the reason I believe is not that I am compelled by the evidence that you are minded but only that implicitly believing serves an evolutionarily useful anticipatory function.

So the traditional PoOM is strongly upheld in two ways. First, I really do (at least implicitly) believe that you are minded and it is reasonable to wonder whether this implicit belief can be justified. It is tempting to say that if by some feature of my psychology I can't help but believe in other minds then I am eminently justified in believing in other minds. After all, what better justification is there for doing something than not having any other option? But this is equivocating on justification. In the epistemological sense, justification means supporting a belief with evidence that it is true. There is thus good reason to hold that my belief is unjustified.

Which leads me to the second way in which the PoOM is upheld: the concept of belief in other minds is dubious by psychological role. Namely, it is evolutionarily useful and psychologically mandated that I believe you are minded, yet I do not have conscious access to the fact that this belief is held on grounds of utility rather than truth. Because of the psychological role that belief in other minds may play, I cannot take my intuition that other minds exist to be in any way indicative of the actual existence of other minds. The PoOM is not dissolved but only strengthened by the possibility that I believe in other minds for evolutionary-based reasons.

The post-Wittgensteinian rebuttal to the PoOM is similarly challenged by mindreading theory. If my understanding of others' behavior proper is being driven by interest in utility rather than truth, then my understanding (especially insofar as it is a token of the so-called lived position) is open to skepticism. I may connect with others in that I interpret their behavior proper to be the same as my behavior proper and thus feel a deep human bond. I may be thoroughly and unfailingly convinced that they are minded in the same way as myself. I am a person and I interpret them to be a person. But I am still interpreting, and moreover likely completely subconsciously interpreting, in order to interact better with them (as opposed to interpreting in order to learn the truth), and there thus always remains the possibility that I am tricking myself. If it is likely an evolutionarily adapted trait of mine that I subconsciously trick myself, then any of my attributions of personhood are dubious by psychological role.

The post-Wittgensteinians may have accurately tracked the concept I use when I attribute a mind to others, and it very well may be the same concept I use to describe my own mind. It does not follow that I am using that concept in accordance with the truth. And if I am not using that concept in accordance with the truth, then the classic skeptical worries that drive the PoOM are stronger than ever. Indeed, it brings me great utility and joy to form bonds with persons, and such bonds rest upon the knowledge that we both are minded, and so why would I ever fail to trick myself when given the opportunity? The more persons the better! To frame the claim more in line with the conceptual problem: especially given the plausible evolutionary origins of the mindreading faculty, does it not still seem likely that I am merely tricking myself into equivocating on the concept of mind? How could our bond rely on anything other than the subconscious mutually acknowledged white lie that we can somehow know each other as minded persons?

My answer is that this line of reasoning misses the point of both the conceptual problem and of Daviesian naturalism. Of course science cannot solve epistemology (or at least cannot satisfy the skeptic with an epistemic claim). The classic epistemological PoOM is hallowed for a reason: it is the very sort of question that science can never answer. There is no scientific evidence that I can imagine convincing the radical solipsist of the existence of other minds. But both the conceptual problem as framed by the post-Wittgensteinians and progressive naturalism as espoused by Davies stress targeting underlying concepts for good reason. The conceptual PoOM is not an attempt to argue against skeptics regarding other minds, but to coerce them to rethink asking the question in the first place. And it is as an aid in this coercion that interdisciplinary work comes in handy.

The above argumentation regarding the utility challenge to the PoOM (and indeed an overwhelming majority of the history of argumentation regarding the PoOM) assumes that other minds are a special kind and then asks how we can know other minds. I, with Strawson, take this to be akin to assuming the earth is flat, and then attempting to calculate the dimensions of the earth. No matter how refined my measurement instruments and no matter how deep my knowledge of geometry, so long as I assume the earth to be flat, my science will not accurately appraise the situation. The task cannot bear fruit, and the skeptic will always prevail. To put the same point another way: there is no way of solving the epistemological problem on its own terms. However, this thesis attempts to show that through solving the conceptual problem of other minds, we can dissolve the epistemological problem. It is in dissolution, rather than solution per se, that science is useful.

When philosophy asks a question, it is not necessarily the job of science to answer that question. Rather, it is the job of the philosopher working with scientific evidence to ascertain whether the question is a silly thing to ask. The scientist will never be able to satisfy the skeptical epistemologist with her account of a human knowing another human's mind. Rather, as we will see in Part III, it is when scientific evidence is used alongside philosophy correctly – as a means of testing dubious concepts – that we obtain meaningful interdisciplinary results. For the progressive naturalist, the role of science is to pave the conceptual path for unhindered philosophical inquiry.

Part III

The Naturalistic Approach to the Conceptual Problem

Part I introduced the two problems of other minds, and explicated the post-Wittgensteinian account. Part II introduced my naturalistic methodology, and explained how this methodology might be fruitfully applied to philosophical debates. Now, in Part III, I do just that. Chapter 5 applies my methodology to the mental transparency assumption, and finds it dubious. This finding means that we lack an adequate account of how we believe in our own minds. Chapter 6 explains how, through the mindreading module, we believe in other minds. Finally, Chapter 7 introduces Peter Carruthers' indirect sensory-access theory of self-knowledge (ISA), which provides the previously lacking account of how we believe in our own minds. ISA takes the mindreading module detailed in Chapter 6 to explain all belief in minds, whether our own or others'.

Chapter 5

One Believes in One's Own Mind

The statement "I believe that I am minded" is foolish and misleading. I don't *believe* I have a mind. I just *have* a mind and know I have it by virtue of having it. Such, anyway, was the Wittgensteinian gist of my dad's response when he saw that my chapter was tentatively titled "I Believe in My Mind". The goal of this chapter is to argue that the position my dad exemplifies errs on the side of conceptual conservatism and is dubious both by descent and by psychological role. Despite the universal intuition of mental transparency, the existence thereof is not in fact a given. It is only by freeing ourselves of the need to premise dubious concepts like mental transparency that we can make progress on the PoOM.

For the post-Wittgensteinians, rejecting the Cartesian viewpoint essentially involves rejecting substance dualism. While rejecting dualism is important to the naturalist cause, it does not go far enough. Descartes' philosophical system hinges on his assumption of infallible access to his own self-presenting mental states. For Descartes, humans undergoing a given mental state necessarily know that they are undergoing that state, and humans who believe they are undergoing a given mental state necessarily are undergoing that state. In his forthcoming book *The Opacity of Mind: An Integrative Theory of Self-Knowledge,* Peter Carruthers notes two facts about the *cogito* that make the extent of Descartes' commitment to this mental transparency assumption clear.

The first is that Descartes' uses of the verb *cogitatio* refer not to a specific reflective kind of thinking but to current propositional attitudes of any kind. Thus *cogito ergo sum* may just as well be translated as "I am in pain, therefore I am" or "I want a snack, therefore I am" as "I think, therefore I am." The point is that Descartes is directly accessing a mental event, not that he is undergoing the particular mental state of reflective philosophical doubt. The second fact is that Descartes does not assume mental transparency without questioning the matter, but rather assumes it because he believes not only that "the claim didn't *require* any argument, but that it *couldn't* be argued for, since it forms one of the basic principles of all knowledge and argument. It was, he thought, as obviously true as anything could possibly be."[31] [32] The idea that he has direct access to mental events is clearly and distinctly true.

At least among philosophers, my dad is far from alone in upholding a weak form of the Cartesian position in 2011. While few philosophers would endorse the mental transparency assumption in as strong a form as Descartes did, there are maybe even fewer who would reject it entirely.[33] Most working epistemologists and philosophers of mind take the existence of mental transparency for granted and jump right to arguing about its implications. Upon conducting an informal survey, Carruthers estimates that at least 95% of philosophers working on self-knowledge of mental states in the last 40 years uphold some version of a weak transparency assumption.[34] Nowadays, this assumption is usually hashed out in terms of privileged access.[35] Privileged access is most importantly a special kind of access; my relationship to my mind is fundamentally different from somebody else's relationship to my mind. While rarely as bold as the raw Cartesian model, all accounts of privileged access preserve some form of a claim to both the infallibility and the self-presentation of at least some instances of one's own mental life.

For example, inner sense theory as advanced by Alvin Goldman (2006) holds that humans have an introspective module that reliably processes and channels one's own

[31] Carruthers 1,16. The manuscript of *The Opacity of Mind* starts over numbering pages with each chapter. Thus, when citing Carruthers throughout this thesis, I will write “Carruthers [Chapter Number], [Page Number].”

[32] Carruthers' italics.

[33] Indeed, even Wittgenstein, a philosopher most would balk at calling Cartesian in any sense, agrees that it is spurious to talk of 'belief' and 'knowledge' with respect to one's own mind. Whereas an other might know of or believe in my pain, I do not know or believe my own pain. I simply have it.

[34] Carruthers 1,14

[35] There is a sense of privileged access that we are obviously warranted in upholding. We have evidence in our own case that we lack in the cases of an other. I can see out of my eyes and hear out of my ears but I cannot see out of your eyes or hear out of your ears. So, in this thesis, I am critiquing privileged access only insofar as it is a form of mental transparency; i.e. the ability to introspect in a Cartesian manner.

mental states to that person just as one's visual system processes and channels visual images. Goldman's view will be explicated as my primary foil in Chapter 7. Another theory, advanced by Richard Moran among others, holds that the expression of the mind constitutes the mind itself. For Moran, humans have access to their own mental states not by introspection, but by avowal: one actively thinks or says "I want a cookie" and thereby commits oneself to wanting a cookie. And, being a rational creature, it is true that one does want a cookie.[36] It is important to note that, while these theories may seem compelling, none of them is perfectly in line with Descartes' clear and distinct intuition. The reason the mental transparency assumption was first articulated in Western philosophy was the powerful intuition that thoughts are directly, necessarily, and immediately self-presenting. This is the same intuition, not of weak and esoteric privileged access but of strong immediate transparency, that I take philosophers who uphold privileged access to have, and indeed that I myself have.

Whether or not Carruthers' 95% figure is accurate, it is safe to say that a vast majority of those theorists who do work directly with the question of knowledge of one's own mind remain committed to mental transparency. They take the scientific evidence about minds on one hand, and the assumption of privileged access on the other, and create theories that reconcile the two. While the way in which we access our own minds may be unintuitive (and intuition is, remember, the reason for assuming mental transparency in the first place), these theorists remain insistent that knowledge of one's own mind is in some sense sometimes direct and non-interpretive, in some sense authoritative, and in some sense sometimes different not just in degree but in kind from

[36] Carruthers 1,16

knowledge of other minds. Those few theorists who completely reject privileged access, most notably Gilbert Ryle[37], are usually radical behaviorists.

This mass assumption of mental transparency is dubious by descent. Descartes' philosophy, despite being groundbreaking, was grounded in theistic ideals. Descartes' other core doctrines, such as substance dualism and infinite human free will, are among the philosophical and theological concepts most called into doubt, if not outright refuted, by progress in science. The preservation of ideals, while very important to Descartes, has no place in 21st century naturalistic philosophy. Indeed, Davies stresses that we should not only avoid making the preservation of ideals a condition of adequacy on our theorizing, but should expect those idealized concepts to be disrupted. When one does make dubious concepts like mental transparency antecedent to argument and analysis, as most philosophers who discuss privileged access do today, one's argument and analysis are fallacious.

Furthermore, the history of the mental transparency assumption shows that the position gets weaker and weaker over time. Since Freud, and exponentially since the rise of cognitive science, most philosophers have buckled under pressure of the evidence and given up claims of Cartesian-strength mental transparency. But even these fervent backpedalers have maintained that access to one's own mind is of a special kind (relative to one's access to other minds). The attempt to preserve historically demarcated special kinds is a hallmark of the conceptual conservatism, and the existence of those special kinds is dubious by descent. Given the unintuitive nature of contemporary theories of access to one's own mind, most philosophers do not attempt to save the intuition of

[37] See Ryle (1949)

mental transparency, but instead only the concept. They do not defend the feeling of transparent access to their own mental states, but only an alienated idea of that access. It thus seems that they attempt to save mental transparency out of conceptual conservatism alone.

Nevertheless, contemporary philosophers may actually have an understandable reason to attempt to preserve the dubious concept that is privileged access. It is the closest they can honestly come to preserving the Cartesian mental transparency intuition and they, along with all other humans, may be compelled to intuit transparency by an inherent feature of their psychology. Humans may have evolved to unquestioningly assume mental transparency. Much like I argued in Chapter 4 regarding the veracity of intuitions of belief in other minds, I will now argue that intuitions of mental transparency are dubious by psychological role.

The mental transparency assumption not only has a strong influence over the course of modern Western philosophy, but across time and place. Carruthers notes the influential role the transparency assumption (usually implicitly) plays in Chinese, Indian, Aztec, and Shuar thought in order to nod towards the possibility that it might be a human universal.[38] More specifically Carruthers presents an intriguing, if far from conclusive, argument from reverse engineering that the transparency assumption might be built into the human mindreading module to serve its own evolutionary function. Carruthers holds that one should expect the evolution of an innate mindreading module (which I will argue for in Chapter 6) to involve the adoption of an innate intuition of strong Cartesian mental transparency. Given the complexity and speed of social interactions, humans need to

[38] Carruthers 1, 5

assume they know something concrete about the situation at hand in order to socialize efficiently. And mental transparency, whether or not it exists, is the simplest and most powerful streamlining assumption humans can make.

To underscore the plausibility that the transparency assumption is innate, take the famous Gazzaniga (1978) commissurotomy experiments. These experiments indicate that humans assume mental transparency even when we know that our brain has been damaged in such a way as to deny access to certain of our own mental events. Commissurotomy (or split-brain) patients have had surgery that separates the two hemispheres of their brain so that they do not interact. The left hemisphere of the brain is where the language faculty is housed, and thus the hemisphere to which split-brain patients have access when speaking. Split-brain patients generally have a very good understanding of their surgery and the fact that there exists mental information in their minds that they cannot consciously access.

In one of Gazzaniga's studies revealing the overriding nature of the transparency assumption, a split-brain patient's right eye's field of vision (and thus left hemisphere) was shown a chicken claw and his left field of vision (right hemisphere) was shown a picture of a snowy scene. His full field of vision was then shown many diverse objects, and the patient was instructed to point at the objects that related to what he had seen before. Predictably, the hand controlled by the left hemisphere chose a chicken and the hand controlled by the right hemisphere chose a snow shovel. The experimenter then asked the patient to explain his choices. The patient confidently and immediately replied that he chose the chicken because of the claw, and the shovel to clean out the chicken coop.

To iterate, the patient knew, and could explain, that his corpus callosum had been severed and that he could not access certain information in his own mind. But he did not, and split-brain patients never did in Gazzaniga's myriad experiments, appeal to his condition and the fact that he might not be able to access the reason he chose the shovel. Plausibly because of a built-in transparency assumption, the patient truly believed that he knew for certain why he had chosen the shovel. In fact he did not know, but was subconsciously interpreting his own action. No wonder it was clear and distinct to Descartes that his mental life was transparent to him: humans might have evolved to unquestioningly believe it. It was clear and distinct to Gazzaniga's patient that he had chosen the shovel in order to clean out the coop, and he was wrong.

This effect is not only present in people with brain damage. Indeed, there is a vast amount of research that shows subjects interpreting their own mental states and then assuming to have transparently introspected them.[39] For example, Wells and Petty (1980) show that people interpret themselves as agreeing with something merely because they happen to be nodding as it is being said. In this classic study, one group of subjects was informed that they were engaging in product testing by nodding their heads as they listened to a tape through headphones. The other group was instructed to shake their heads while listening to the same tape. When questioned afterward, members of the nodding group were significantly more likely to believe the things said on the tape than were members of the headshaking group.

[39] Some more of this literature will be examined in Chapter 7. For yet more, see Wilson (2002), Wegner (2002), and Carruthers Chapter 10.

Despite knowing that they were nodding or headshaking only in order to follow instructions, the subjects unconsciously took their own nodding or headshaking as evidence that they had a strong stance on what they heard. A number of control experiments have been done to rule out other explanations, such as the possibility that nodding simply makes one more agreeable.[40] In fact, we interpret ourselves nodding in exactly the same way as we interpret others nodding, even when we are explicitly aware that we are nodding for a different (and trivial) reason. And, once again, it is clear and distinct to us that our interpretation is in fact an introspection, and we are dead wrong.

I have by no means produced enough evidence to warrant renouncing the transparency assumption. Some version of an inner sense or other theory of knowledge of one's own mental states may turn out to be true, and the dubious concept of privileged access may very well be saved. However, especially given the extreme conceptual conservative faith usually implicitly upholding mental transparency, I do believe I have at least given enough reason to doubt that it is necessarily the case. I have also given grounds that one plausible alternate hypothesis to consider is that mental transparency is best explained as an illusion produced by a mechanism of our psychology.

The concept of privileged access is dubious by descent and the intuition of mental transparency is dubious by psychological role. Hence, in line with Davies' naturalist directives, we should bracket the concept of mental transparency and explore what we actually do know about our own access to our own mind before theorizing about that access. And then that theorizing should not take mental transparency to be a premise, but as one of several possible conclusions. There is a belief about my own mind, as about

[40] See Briñol and Petty (2003) and Carruthers 10,16

other minds, that needs justification. Indeed, it is plausible that, whether by force of social indoctrination or evolutionary psychological function, I cannot help but believe in my own mind. Justification of that belief with regards to truth is thus all the more important, and while that justification may well end up being mental transparency, it should not be assumed that it is so.

If our best science does not point towards the existence of mental transparency then, because we are not assuming it to exist, we should not take it to exist. In the remainder of Part III I defend the claim that Carruthers' Indirect Sensory-Access Theory of Self-Knowledge (ISA) is our best scientific theory of self-knowledge. ISA does not take mental transparency to exist, and thus, I will argue, neither should we.

Chapter 6

One Believes in Other Minds

As touched on earlier, mindreading is the human capacity to attribute mindedness. The two main strands of mindreading theory are called 'theory theory' and 'simulation theory'. Theory theorists hold that children are like little scientists who, through trial and error, construct a scientific theory about the existence of other minds which they actively use in attributing mental states. Simulation theorists, on the other hand, hold that humans understand other minds by learning to simulate others' minds in their own minds. They take what they know about a given situation and, barring any conflicting information they know the other to possess, simply attribute their own mental processes to the other. Both theory theory and simulation theory require a great deal of (and thus time spent) learning before mindreading can become operational. Theory theorists must learn to whom it is appropriate to attribute what sorts of mental states, and simulation theorists must learn whom to simulate.

While there is a huge amount of literature on the theory/simulation debate, I will not go into detail regarding the merits of theory theory versus simulation theory because in the end I will take some understanding of each theory to be correct. What follows is a survey of core mindreading research which inductively supports a third account of mindreading: that humans possess an innate mindreading module. A module, in cognitive science, is "a specialized function-specific processing system with its own specific neural realization."[41] Thus, my position[42] is that the human mind has evolved a specialized systemic capacity to, through both learning and simulation, formulate belief in and understanding of mental states.[43]

The most widely accepted claim regarding the mindreading capacity is that normally-functioning humans acquire an understanding of other minds around the age of four. This claim is supported by abundant research on children's ability to pass false-belief and misleading-appearance tasks. For example, in a much-discussed experiment by Wimmer and Perner (1983), experimenters act out a story for children. In the story, Sally places a marble in a basket. Sally then leaves, and while she is gone Anne moves the marble to a box. Children are asked either where Sally will look for her marble when she returns, or just where Sally thinks her marble is. In both cases normally developing four year old children typically pass such tasks and point towards or verbally indicate the basket. Three year olds, along with much older autistic children, typically fail the task and point towards or verbally indicate the box. These results have been replicated and confirmed across experimental paradigms with tasks testing varied false beliefs among children from varied countries, cultures, and ethnic groups, and as such are almost always taken as canon by theory, simulation, and module theorists alike.[44] It is just about as close to known as scientifically possible that humans, from the age of 4 on, believe in other

[41] Carruthers 7, 5

[42] Informed by the modular mindreading accounts put forward by Leslie (1994 and 2004) and Baron-Cohen (1995).

[43] Leading modular accounts of mindreading, including the one I espouse in this thesis, actually posit a 2 system mindreading faculty. Humans engage in both immediate intuitive mindreading, of the unconscious sort discussed extensively in this thesis, and slower reflective mindreading, which might actually be a conscious process. For the sake of brevity I do not go into 2 system mindreading in this thesis, but for a full account of its implications for (and the support it lends to) Carruthers' theory see Carruthers Chapter 7.

[44] Leslie (2004) 528

minds.

Nonetheless, a new research paradigm has begun to produce results that strongly suggest that children much younger than four years old have the ability to discern false belief in others. Rene Baillargeon notes that traditional investigations into children's false-belief understanding, such as the Sally-Anne experiment, rely on elicited-response tasks in which the children verbally respond to verbal prompts which reveal their understanding of the mental states at play.[45] The new paradigm, on the other hand, uses what Baillargeon terms spontaneous-response tasks rather than elicited-response tasks. These tasks do not depend on children's explicit responses to experimenter's verbal prompts, but rather study children's understanding of false belief by noting the behaviors with which the children spontaneously respond to a situation. One particularly well supported type of spontaneous-response task, the violation-of-expectation task, studies whether infants gaze longer when they observe actions incompatible with actors' false beliefs.

For example, research by Surian et alia (2007) shows that 13-month-old infants use their mindreading capacities to attribute false belief about the location of an object, and attribute belief to caterpillars as well as to humans. In this study, there were two screens, one hiding an apple and the other hiding a wedge of cheese. The infants first witnessed several familiarization trials in which the caterpillar watched the experimenter hide the foods and then consistently went for its preferred treat. Then, in the experimental trial, the foods were hidden in opposite locations before the caterpillar entered the scene. The infants gazed significantly and reliably longer when the caterpillar, whom they had

[45] Baillargeon et alia 110

witnessed to prefer one type of food over the other, did not return to the original location of the preferred food. The infants mistakenly attributed a false-belief to the bug.

Another type of spontaneous-response task, the anticipatory-looking task, studies whether an infant's gaze anticipates where an actor with a false belief about the location of an object will look for that object. For example, Southgate et alia (2007) used a non-verbal anticipatory-looking task to discover whether 25-month-olds anticipate the actions of a person with a false belief. In familiarization trials, a bear puppet hid a toy in one of two boxes while the actor looked on, and the actor then opened a door corresponding with the correct box from which to retrieve the toy. In the experimental trial, the actor once again looked on while the bear puppet hid the toy in a box, but, after the toy was hidden, was distracted by a ringing telephone and looked away. While the actor was looking away, the bear puppet removed the toy from the box and left the stage with it. When the phone ceased to ring and the actor turned back to the stage, the infant correctly anticipated the actor's behavior and gazed at the box where the infant took the actor to falsely believe the toy was hidden.

These studies, along with the many others in this ever-growing research paradigm[46], indicate that children have an active mindreading capacity by not the fourth, but the second year of life, if not much earlier. And the earlier in human life the mindreading faculty is found to be active the more support there is for a modular account, because both theory theory and simulation theory require a significant amount of learning time before they can take children to be able to mindread. But before any claims about

[46] See Baillargeon et alia (2010), Onishi (2005 and 2007), Scott (2009 and in press), and Song (2008a and 2008b) for yet more spontaneous-response studies.

the ramifications of infant mindreading can be persuasive, there is an obvious concern to address. Namely, why is there such a large gap between ability to perform in spontaneous-response and elicited-response tasks? Why, if one can mindread by the 13th month of life, does it take that same child until the fourth year to report that ability?

Baillargeon et alia (2010) propose what they term the "response account" to describe why children who can mindread nonetheless fail elicited-response mindreading tasks. The response account states that elicited-response tasks require people to carry out three distinct processes. The child must first engage in false-belief-representation. She must understand the false-belief without conflating it with her own belief. Secondly, the child must engage in response-selection and access the understood false-belief in order to select her response. She must select, from among all possible responses, that response which takes into account the facts that she understands the false-belief and is answering as if she were the other. Finally, the child must engage in response-inhibition and inhibit any urge to respond with her own belief rather than the false-belief of the other.[47]

In spontaneous-response tasks, on the other hand, infants must only engage in false-belief-representation and they automatically behave in such a way as to spontaneously give away their understanding of the false belief. The response account thus argues that, until the age of four, children are overwhelmed by the complexity of the multi-step elicited-response requirements. Baillargeon reports that findings in neuroscience support this account, suggesting "that (1) the right temporo-parietal junction plays an important role in the false-belief-representation process, (2) regions of the anterior cingulate and prefrontal cortex play an important role in the response-selection

[47] Baillargeon et alia 114

process, and (3) the connections between the frontal and temporal brain regions mature later and more slowly than other connections." The response-selection process has difficulty accessing the false-belief-representation process in the immature toddler brain, and children younger than four thus predictably pass spontaneous tasks but fail response tasks.[48]

Carruthers introduces another factor that helps explain the gap between mindreading and the report thereof: epistemic vigilance. As an infant, a child has no need for a high degree of epistemic vigilance in communication, as she is likely to only interact with primary caregivers, and primary caregivers are highly unlikely to harmfully lie to her.[49] Indeed, in order to learn something as multifaceted as a native language at so early an age the child must take everything at facevalue, or else get lost in complexity. And in an insulated loving community false-belief is a conversational nonstarter. It is only once the child's social circle starts to broaden and she interacts with other children and less trustworthy adults that falsehood gains an important role in conversation. Once this happens, the child's epistemic vigilance rapidly rises, and around the age of four she is fully linguistically equipped to competently discuss false belief. Hence late-signing deaf children experience a delay in acquiring the ability to perform in elicited-response tasks because they interact with non-caregivers later than normal children.[50] Likewise, children with older siblings acquire an early ability to perform in elicited-response tasks because they cannot trust their brothers and sisters from the start.[51] All of these children are equipped with a working mindreading module from at least 13 months of age, if not

[48] Baillargeon et alia 115

[49] Carruthers 9, 27

[50] Carruthers 9, 28

[51] *ibid*

from birth, but they start to use it in active communication at varied times based on their individual social situation.

False-belief understanding in infants, in addition to making an innate modular account of mindreading plausible, also makes it unlikely that mindreading is housed in the language module. Theorists such as Gordon (1986) and de Villiers (2007) have long argued, from both simulation theory and theory theory standpoints, that mindreading is simply an offshoot of language acquisition. But infants can both mindread before they can speak and speak without being able to report mindreading. While the development of language might play a role in being able to report mindreading by the age of 4, Baillargeon's work indicates that the mindreading module develops earlier than the language module. Varley's (1998 and 2001) agrammatic aphasia studies underscore this objection to the claim that mindreading and language are produced by a singular module. These studies show that patients who have undergone severe left brain hemisphere damage and have lost almost all linguistic competence nonetheless pass false-belief tasks. So the two modules can operate distinctly, at the very least.[52]

If mindreading is indeed modular then, as the language module can be damaged or absent, the mindreading module should similarly be damaged or absent in certain humans. A great deal of work on the performance of autistic people on mindreading tasks suggests that autism might be what constitutes just such damage or absence. Simon Baron-Cohen goes so far as to claim that the behavioral impairments that characterize autism can all be traced to impairment of the mindreading system. Baron-Cohen terms

[52] Carruthers 7, 3

the autistic person's lack of a capacity to mindread "mindblindness."[53]

Numerous studies have shown that autistic children are unable to pass elicited-response false-belief tasks until the age of nine, and often are unable to pass more subtle mindreading tasks, such as understanding irony, their entire lives.[54] Furthermore, once able to pass mindreading tasks, there is ample evidence that autistic people do so not by mindreading but by following heuristics they have devised in order to get by without a mindreading faculty.[55] For example, Senju et alia (2009) showed that adult autistic people fail the same anticipatory looking tasks that the bear puppet experiment proved two year old normal functioning humans pass. Without being specifically prompted to respond appropriately, the autistic person does not have the opportunity to employ the makeshift heuristics she plausibly uses to make up for her mindblindness.[56] This evidence is not quite strong enough to be sure that what makes autistic people autistic is a lack of or impaired mindreading module. But there is, as far as I am aware, nothing in the evidence that runs counter to the claim.

There is also a large amount of experimental data on the mindreading capacities of chimpanzees and other primates. If an evolutionary modular account of mindreading is right, then the evidence should show that chimpanzees (along with other highly-social primates and even some types of birds) have a more rudimentary but similarly functioning mindreading module. Call and Tomasello (2008), upon reviewing the 30 years of research on chimpanzee mindreading since Premack and Woodruff's seminal 1978 "Does the chimpanzee have a theory of mind?", find that the evidence is in line with

[53] Baron-Cohen 1

[54] Carruthers 9, 9

[55] Perhaps theory theory holds true for people lacking modular mindreading ability.

[56] Carruthers 9, 10.

just such a conclusion. In particular, while chimpanzees consistently fail false-belief tasks, they do nonetheless engage in a more limited form of mindreading.

Chimps, like humans, come to understand others not just in terms of behavior but in terms of the goals, intentions, perceptions and knowledge underlying that behavior. When competing with others for food, the apes take into account a whole range of mental states. For example, Hare and Tomasello (2006) had chimpanzees compete for food with a human inside a glass booth. The chimps chose to approach the food along the side of the booth that featured an opaque barrier blocking the human's line of vision. In an extension of that experiment conducted by Melis et alia (2006), the chimps accessed the food by choosing to go through a tunnel that featured a more or less silent door over choosing a tunnel that featured a noisy door. Chimps understand what their fellow foodmongers can and cannot see and hear and, when it is to their advantage to do so, act in order to trick others.

The particularly strong capacities of the human mindreading system, and the ways in which humans use mindreading in complex social interactions, lead some researchers to hold that human mindreading is distinct from mindreading in other animals in that it is centered on empathy. Following the research of Tomasello among others, anthropologist and primatologist Sarah Hrdy holds that the key to the principle difference between humans and other apes lies in the mindreading module. Despite nearly identical cognitive abilities, humans, when compared with other apes, simply pay much more attention to what other individuals are thinking and feeling, and, perhaps even more tellingly, are much more eager to share and cooperate with other individuals.

Hrdy's 2009 book *Mothers and Others: The Evolutionary Origins of Mutual Understanding* espouses an account of evolved human nature that takes empathy to be central to humanness. She advances the argument that the mindreading module first evolved during Pleistocene era, around the same time as the emergence of the first species under the genus *homo*. Human babies take a very long time to become self-sufficient, and their hunting-gathering parents often could not provide for and protect them on their own. Thus, supplementing parenting with regular alloparenting – the act of parenting by non-biological parents – became the norm among very early humans.[57] And alloparenting only occurred because humans were able to see other humans as minded, and thus worth protecting. The very earliest hunting and gathering societies only successfully produced large, vulnerable, and slow-maturing offspring by virtue of empathic alloparental care made possible by the mindreading module. So, according to Hrdy, thousands of years before the advent of anatomically- and emotionally-modern humans, a particular empathic mindreading capacity distinguished early members of the *homo* genus from other apes.

A widespread anthropological view is that what sets humans apart from other animals is the language faculty. Hrdy posits that human specialness also, or even instead, might be due to a more advanced capacity for mindreading.[58] The Baillargeon studies certainly indicate that language and mindreading modules are distinct, and thus that the specialness of human mindreading is not merely a byproduct of the specialness of human language. Regardless, the evolutionary advantages of mindreading, and in particular of

[57] Hrdy 176

[58] Hrdy 3

mindreading being innate and modular in nature, are obvious. It is immensely plausible that humankind developed the ability to mindread primarily because it was useful for forming tightknit extended families which made the rearing of slow-maturing children possible.

To summarize, the best inductive conclusion to draw from the abundant evidence on the mindreading capacity is that it is an evolved innate module by which humans subconsciously assign mental states to others. And honestly assigning mental states to others is only humanly possible if one believes in the minds of those others. Thus, the mindreading module is the biological impetus for belief in other minds. The problem of other minds is the problem of justifying those beliefs that are produced by the mindreading module.

In Chapter 5 I determined that we believe in our own minds, and that this belief (because the mental transparency assumption is dubious) is one that needs justification. In this chapter I explained how we believe in other minds, and noted that this belief is what the PoOM claims needs justification. In Chapter 7 I will tie Chapters 5 and 6 together and argue that how we believe in other minds is exactly how we believe in our own minds. In Part IV I will argue that the implication of this fact for the PoOM is that our belief in other minds requires no differential justification than our belief in our own minds.

Chapter 7

One Believes that Others are Minded in the Same Way as Oneself

The above chapter title is stated ambiguously in order to point out that philosophers dealing with the conceptual problem have consistently equivocated on two very different meanings of the statement. Put this way, the title could be taken, as the conceptual problem is traditionally taken, to mean that the concept of mind attributed to oneself is the same as the concept of mind attributed to others. This is P3 in the argument that motivates the PoOM from Chapter 2 . But it could also mean that the way by which one believes oneself to be minded is the same as the way by which one believes others to be minded (this is what, for the post-Wittgensteinians, undermines P5). The latter is what the post-Wittgensteinians argue, while at times claiming to argue for the former. But showing the latter does not necessarily amount to showing the former. After all, I might attribute mindedness to a cat or robot using the same mindreading methods I use to attribute mindedness to a human adult, but it is not clear that I am attributing the same concept of mind to each subject.

Thus, it falls to my analysis of the relevant sciences not only to support the post-Wittgensteinian claim that the way in which minds are perceived is univocal, but also that the perceived minds are themselves conceptualized univocally. But I am getting ahead of myself. In this chapter I will argue that the psychological and neuroscientific evidence strongly suggests that the way by which one believes – namely the psychological mechanism that is mindreading – is univocal. In Chapter 8 I will address whether knowing that the mechanisms that prompt belief in one's own mind and others' minds are one and the same is sufficient evidence with which to solve the true conceptual problem of other minds.

A claim regarding the equivalence of the mechanisms by which one believes in the two “kinds” of mind (mine and others') must obviously start with what is known about the mechanism at work in each kind of belief. In Chapter 6 I explained that the mindreading module is the mechanism by which we believe in other minds. I also explained, in Chapter 5, that the transparency assumption is not necessarily a good account of the way in which we believe in our own minds. Thus, it now falls to me to explain an acceptable account of the mechanism by which we believe in our own minds which does not premise the transparency assumption. In doing so I will explicate what I take to be the most thorough and convincing account of knowledge of one's own mind: Carruthers' indirect sensory-access theory (ISA). ISA takes transparent access to propositional attitudes to not exist and the module by which one believes in one's own mind to be the mindreading module itself. Upon examining the relevant scientific evidence, I will conclude that Carruthers' is the altogether best available theory and that it is thus plausible that the way by which one believes oneself to be minded is the exact same as the way by which one believes others to be minded.

Alvin Goldman (2006) is Carruthers' main competitor with regards to theories of self-knowledge. Goldman's simulationist inner sense theory takes transparent access to propositional attitudes to exist. For Goldman, humans possess a special channel of introspective access to their own mental states. Moreover, as Goldman is a simulation theorist, he holds that attributions of mental states to others are also grounded in introspection. Humans use their introspective module to simulate others' perspectives and mental states. As Carruthers notes, for Goldman "the final step in each episode of mindreading is to identify the mental state in oneself with which the simulative process has concluded, and then to attribute that mental state to the other person."[59] Because one must be capable of self-knowledge before simulating other-knowledge, Goldman must maintain that the human capacity for self-knowledge evolved before the capacity for other-knowledge, and that it must emerge significantly earlier in development. As we will see, Carruthers challenges both of these points.

On the positive side of things, Carruthers' ISA theory comprises two claims. Firstly, the human mind contains a single faculty that is responsible for the attribution of all mental states. The mindreading faculty, by means of interpretation, is the source of self-knowledge as well as other-knowledge. Secondly, the inputs to this mindreading faculty are all, broadly construed[60], perceptual in nature. Some sensory states which constitute perception are thus introspective and non-interpretive, but the vast majority of mental states, including propositional attitudes of the sort detected in mindreading tasks, are second-hand and interpretive. I will now further explicate and give evidence to substantiate both of these claims.

The first claim relies heavily on arguments from reverse engineering. As I noted in Chapter 6, there are good evolutionary reasons for human beings to have evolved a mindreading module. Effective communication with and empathy for fellow humans

[59] Carruthers 6, 6

[60] Broadly construed in that perception here includes proprioception (the sense of the relative position of neighboring parts of one's body) and interoception (sense that is stimulated from within one's body) along with visual imagery and inner speech.

greatly increases one's fitness and ability to produce and raise offspring, and this held especially true in hunter-gatherer societies. Sarah Hrdy shows that there is good evidence that 3rd person mindreading evolved very early in human development, probably during the early Pleistocene.[61]

Conversely, there is no evolutionary reason, or at least no reason that seems powerful enough to be selected for, for a standalone introspective inner sense module to have evolved before the ability to mindread others. Goldman certainly gives no such reason. Rather, it is much more likely that the ability to express one's own mental states is, albeit an arguably very important adaptation in human evolutionary history, merely an afterthought of a mindreading capacity originally geared towards other minds. Furthermore, alongside its redundancy with the mindreading faculty, there are well-supported neuroscientific considerations that count against the likelihood that humans evolved an introspective faculty.

For one, there is the simple fact that neural connections take a lot of energy and nutrients to build and maintain.[62] The brain only has so many resources to go around, and it does not make much sense to devote a great deal of them to introspection when co-option of the already evolved mindreading module is nearly as effective at self-ascribing mental states. A second consideration is that non-sensory conceptual modules are physically located all over the place in the brain, rather than in one spot from which they could all be pipelined to consciousness.[63] An introspective faculty would require its own complicated and expensive physical manifestation in the brain. And there would have to

[61] Hrdy 283

[62] See Aiello and Wheeler (1995)

[63] Carruthers 2, 16

be one or more expensive physical channels connecting the inner sense module to the conceptual systems dispersed around the brain (from the frontal lobes that form judgments, to the premoter cortices that form intentions, et cetera). There would hence have to be very compelling evolutionary pressures to evolve an inner sense module. And leading inner sense theorists have not, to date, offered accounts of any such pressures.

Much less taxing, and thus much more likely to have evolved under weak evolutionary pressures, is the presence of an innate transparency assumption built into the 3rd person mindreading capacity (as argued for in Chapter 5). With the intuition of transparency assumption in place, 1st person mindreading of the sort ISA predicts only requires turning one's mindreading attention and interpretation on oneself. While an argument towards evolutionary parsimony such as this one is never completely persuasive, it is quite compelling until (if ever) relevant evidence against it comes to light. No such evidence – in this case concrete findings of anything in neuroscience or psychology which highlights a distinct introspective capacity – is on offer. That the appeal to evolutionary parsimony integrates with current work in neuroscience bolsters its credibility.

One prevalent objection to ISA is that I am much more confident in dealing with my own propositional attitudes than those of others. I do not face the sorts of ambiguities and confusions that arise in communication with others when thinking about my own statements. And I often feel like I am involved in interpretation when ascribing mental states to others, but rarely when ascribing them to myself. But the ISA account should not find these claims surprising. That I use one faculty to mindread both myself and others is not to deny that I am better at mindreading myself than others. The process being the same does not entail equally sound results when the inputs can differ so substantially. After all, the mindreading faculty has a great deal more evidence at hand when attributing mental states to myself. Visual imagery, previous inner speech (i.e. verbal thought), and bodily feelings are all modes of perception that aid in self-mindreading and not in other-mindreading.[64] But note that they aid, as in help interpret, as opposed to constitute the mindreading. I do not simply come to know my own mind by way of the ample evidence I have access to (that I do not have about other minds), but rather just more accurately interpret what might be going on in my own mind. I should be more confident in my conclusions, as I have more to work with.

Moreover, especially in conjunction with the built-in transparency assumption, self-interpretation is often self-fulfilling. I can misinterpret my own belief, but that misinterpretation readily becomes my new belief (and subsequent evocations of that belief are actually correct interpretations). I am under the thrall of the transparency assumption, and as such often take my interpretation of my belief as gospel. I am thoroughly convinced that my interpretation was introspection, and therefore henceforth will really believe that which I formerly may or may not have really believed. Gazzaniga's patient did not select the shovel to clean the chicken coop, but he does now really believe that he selected it for that reason, and if shown a chicken coop in the future his true object of choice may well be a shovel. So, I am confident in my interpretations of

[64] It might be argued that this counts as privileged access. And it does, in the broad sense: I have access to evidence no one else does. But it does not count as Cartesian privileged access in that it is not in any way at any time infallible nor self-presenting to consciousness. It is not the sort of privileged access that vindicates the mental transparency assumption and drives the classic epistemological problem of other minds.

myself not only because they are well founded, but because I am not even aware they are interpretations. The fact that I generally do not doubt my understanding of my own beliefs only goes to show that I, like Descartes, am highly prone to assume I have transparent access to my mental states. It does not show that I do in fact have transparent access.

There probably, understandably, remains some confusion regarding the distinction between what may or may not have access, and what that thing may or may not have access to. If I interpret my own mind, and especially if I often err when doing so, what exactly differentiates the 'I' that interprets and the 'I' that is interpreted? Are my conscious self and my nonconscious self somehow, like Stevenson's Jekyll and Hyde, different beings altogether? These questions lead me to the second claim of ISA: that the inputs to the mindreading module are all perceptual in nature. To elucidate this matter, Carruthers appeals to Bernard Baars' (1989) global broadcast model of the mind. Global workspace theory, as proposed and defended by Baars, has been widely accepted as an accurate account of mental processes and access consciousness.[65]

Baars bases his global broadcast model on the blackboard analogy popular in artificial intelligence research. The basic idea of the model is that the mind is organized around a common workspace and that information from this workspace is globally broadcast to all the various processing systems throughout the brain. Imagine specialists in all of the domains of knowledge gathering in a room with a blackboard and contributing information to that blackboard without directly interacting with one another. The specialists, on this analogy, are all of the sensory systems of the mind: vision,

[65] See Shanahan (2006), Dehaene (2001 and 2007), Baars (2010), and Carruthers Chapter 2.

hearing, touch, taste, smell, bodily feeling, visual imagery, inner speech, and so on. These various means of perception globally broadcast the information they have processed (so that, for example, the language module has access to what the body has heard so that it can do its job). The blackboard is the site of consciousness. For Baars, "all and only the contents of global broadcasts are conscious, and the fact of their being broadcast *explains* their conscious status."[66] [67]

There is no higher-order executive system that controls who writes what on the blackboard. Rather, the specialists independently compete for access[68] to the means to globally broadcast their ideas, and their information is filtered only by themselves. The specialists write what they have gleaned from their research – thus, as vision science confirms, the visual system does not simply relay photographic images to the mind but rather processes and interprets visual information and relays a heavily doctored account of its inputs[69] – on the board, and in doing so make information available both to consciousness and to various cognitive processing systems in the brain. This writing on the board constitutes low-level conscious awareness.

The conscious mental event (the token writing on the board) also acts as an input for a number of nonconscious modules, including conceptual, affective, and executive systems. These systems might too be represented by specialists in the room, but they do not directly contribute information to the blackboard. Rather, they read the blackboard, use the information there for their own purposes, and then tell their findings to the original specialists. The original specialists, meanwhile, while continuing to write their

[66] Carruthers 2, 2

[67] Carruthers' italics

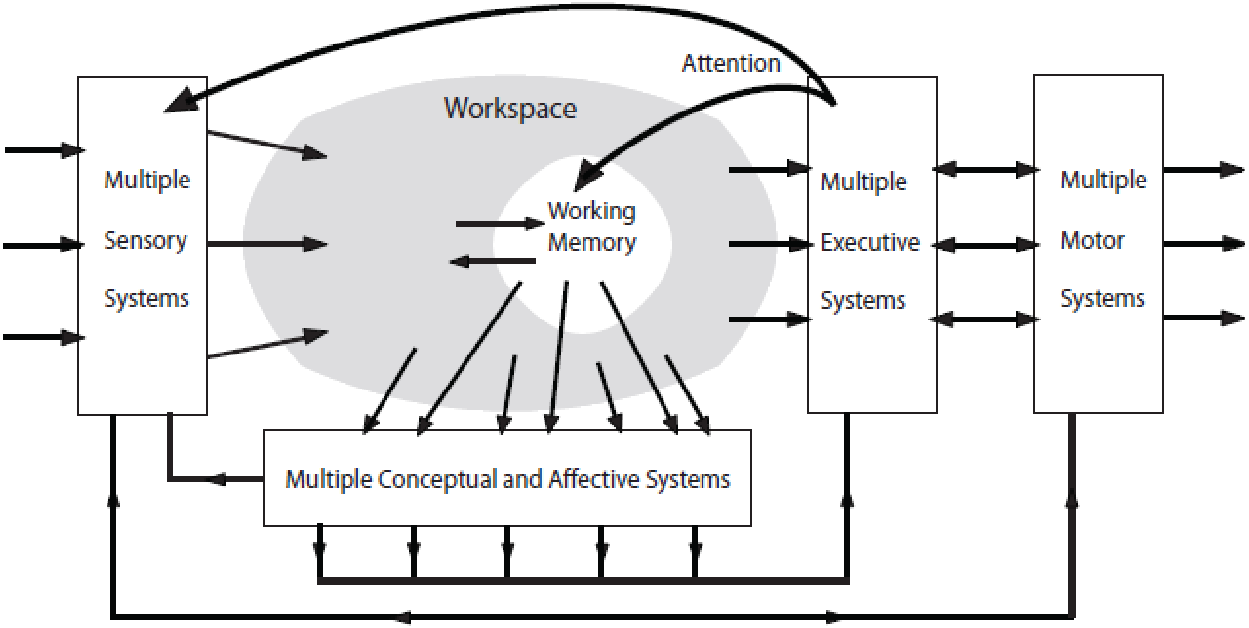
[68] For an account of this competition, see Desimone and Duncan (1995)

[69] See Nijhawan (1997)

own findings on the board, also jot down compelling information from these secondary specialists. So not only are the perceptual systems the only parts of the mind that directly feed consciousness, but other systems act on consciousness only by going through perception (usually in the form of visual images or inner speech). Perceptual information fed into the mind by sensory systems produces a conscious event and a global broadcast to the various processing centers of the mind. This information sparks nonconscious processing in conceptual systems that feeds back into sensory systems and issues in further conscious events. This process continues ad infinitum and constitutes the so called stream of consciousness.

The mindreading module makes sense as a second-level conceptual faculty. The mindreader is a specialist who does not herself write on the blackboard, but who interprets information on the blackboard and whispers her interpretation into the ears of the scribe specialists. Hence, as ISA entails, all of the inputs to the mindreading faculty are perceptual in nature: they are all provided by the original set of specialists. Further, the information uncovered by the mindreading faculty, for example the belief that Sally will mistakenly look for the marble in the basket, is only broadcast in sensory clothing. So, I do not actively consciously believe that Sally will look in the basket, but see that she last saw the marble in the basket, which my mindreading module interprets as believing that Sally will look in the basket. My mindreading module then tells my inner speech faculty to broadcast the feeling that I believe that Sally will look in the basket. This feeling is literally an object: a token image of, for example, the inner speech fragment 'Sally will look in the basket.' Images make use of the same cognitive mechanisms as perception, and are thus globally broadcast in the same exact manner.[70] All conscious beliefs about the mental states of others (and ourselves) are in fact only images – processed perceptions – of the unconscious beliefs produced by the mindreading module.

FIGURE A[71]: The Global Broadcast System.



Sensory information from the eyes, ears, et cetera is fed into sensory systems, which process, package, and relay information to the global broadcast workspace and consciousness. Notice that the only systems that feed into global broadcast (the greyed area labeled "Workspace") are sensory systems. Global broadcast then distributes the sensory information to various executive, conceptual and affective systems throughout the mind. The mindreading module is one of these conceptual systems. These systems process the information and relay it further around the mind. However, before the conceptually processed information is ever relayed back into global broadcast (and thus consciousness), it is always sent back through the sensory systems. The sensory systems package the information into perceptual form, and feed the packaged information into global broadcast and consciousness, starting the cycle over again.

[70] See Kosslyn (1994)

[71] Adapted from Carruthers Figure 2.1

Because all of the conceptual information produced by the mindreading module is processed by sensory systems before reaching consciousness, it is only this processed perception of a feeling that I believe, desire or intend something – rather than a 'pure' belief, desire or intention – that is accessible to consciousness. To clarify this point, remember the visual system once again. My vision is not photographic; not simply an exposure of the latent image captured from the reflection of light. My visual images are not real-time reproductions of what my eyes see. Rather, my visual systems process what my eyes see, adapt that information into images that are useful for navigating the situation at hand, and then globally broadcast these preprocessed images.[72] Inner images of propositional attitudes are preprocessed by my sensory systems in the exact same way. Davies notes that "naive realism concerning our vision ... is out of the question."[73] So too is naïve realism concerning our own propositional attitudes.

Consciousness is thus, in a sense, passive. It simply is the relaying of nonconsciously preprocessed information. I do not consciously do anything; consciousness is simply an awareness of the things I do. For another example, as I type I do not consciously believe I am wearing pants. I sense that I am wearing pants and this sensation, after being processed by my sensory systems, is globally broadcast.. And my mindreading system interprets this broadcast sensory information and leads me to believe I am wearing pants. But I do not have direct conscious access to this belief. Instead, the belief is fed back through and processed by my sensory systems before reaching consciousness. Consciously, I have only the sensation of believing that I am wearing

[72] Davies 118
[73] Davies 119

pants.

Notice, here, the likelihood that concepts dubious by psychological role arise from the global broadcast system. I do not consciously stride forth and define my concepts such that they accord with the truth as best as I can fathom it. I do not even believe that I am wearing pants because of a conscious review of the evidence. Rather, I am simply imagistically informed by my unconscious mind that I believe I am wearing pants, or that I believe that I am the conceptual center of the universe. And, because all of my cognitive faculties that serve as various inputs to the global broadcast system are evolved modules, I have very little ground for holding that they point me towards the truth. Rather, they point me towards what is evolutionarily most useful for me to consciously think I believe. Thus, every concept that they do point me toward, no matter how powerfully they point, is likely dubious by psychological role.

All information that is globally broadcast is sensory, as consciousness is purely sensory. This is not to say that all perception is globally broadcast. Not all, or indeed most, information available to the specialists is written on the blackboard. Rather, some sensation never reaches conscious awareness yet nonetheless affects nonconscious modules. This fact has been amply proven by priming experiments in psychology. Take the famous Bargh, Chen, and Burrows experiment (1996) in which subjects were ostensibly in the lab to complete a language task. Half of these subjects were given a language task, and the other half were given the same task with keywords swapped out for words clearly pertaining to the elderly. Upon completing the task, the subjects were told they were free to go, and walked out of the lab, down the hallway, and to the elevator which led to the exit. A confederate of the experimenter sat in the hallway, pretending to be the next subject in line, and surreptitiously timed how long it took each subject to reach the elevator. Astoundingly, those subjects who had completed the task with words pertaining to elderly people walked significantly slower than the students in the control group. In the face of this study, and the many like it, it is obvious that our perceptual systems affect our actions in ways which are not consciously accessible. The consequences of this fact for the PoOM will be noted in Chapter 10.

In addition, much of the perceptual information broadcast in consciousness is pre-conceptualized. Indeed, if perceptual systems did not interact with conceptual systems before global broadcast took place, they would fail to exclusively broadcast that information which is worthy of attention. So, some of the specialists take each other aside and give information before even entering the room with the blackboard. Because of this pre-blackboard collaboration, globally broadcast inputs, while wholly sensory in nature, do carry imbedded conceptual information. Otherwise, if first broadcasts were purely perceptual, we would not immediately see things as unitary wholes but as chunks of color and shape. We actually do see things categorized as unitary books, faces, and apples.[74] Even our most immediate perceptual judgments are already laced with conceptual content. The upshot of this for the mindreading faculty is that the way in which one interprets minds will have already been informed and biased before conscious global broadcast occurs.[75]

The evolutionary benefits of a global broadcast system are significant. For one, it

[74] Carruthers 2, 25

[75] While tangential to the goals of this thesis, any defense of the global broadcast system must account for scientific knowledge about the working memory system. Carruthers Chapter 2 does this nicely, arguing that working memory is actually embedded in global broadcast.

avoids unnecessary expenses of the same sort that counted against the evolutionary plausibility of an inner sense module. Moreover, because it consists of blanket-coverage rather than custom pipelined channels of access, global broadcast of perception allows varied structures in the brain to use sensory information without having to maintain direct communication with perceptual systems.

Another strong evolutionary benefit of the global broadcast system is that it easily welcomes new adaptations. A newly evolved cognitive processing module, as the mindreading module was at one time, does not have to come equipped with a network of channels for inputs and outputs. Rather, any new system already comes with the default setting of being attuned to global broadcasts of all the varied forms of perception, which can take care of its core processes while other more refined pre-global broadcast modes of communication can be created over time. Especially given the complex social situations humans rely on such cognitive modules to navigate, it is quite lucrative to allow those faculties access to a wide range of perceptual evidence rather than just that evidence which is pre-judged to be useful and therefore directly pipelined in. More fundamentally, a global broadcast system allows new modules, like computer software, to be added or removed from the architecture of the mind without disrupting the rest of the machine.

Thus far in the chapter I have detailed the core claims of ISA and argued that these claims, perhaps unlike the claims of competing theories, are compatible with an evolutionary picture of human development. In order to fully support Carruthers' theory, though, I must go one step further, and check the validity of its predictions. A good theory must be checked against the world it intends to describe.

ISA makes two main predictions that differentiate it from the other prominent accounts. The first is that disassociation between one's capacity to ascribe mental states to oneself and to others should not exist. If a person can self-ascribe then she should be able to ascribe to others, and vice versa. The second prediction is that, via self-interpretation, one should frequently confabulate sensory evidence and commit mistakes in the ascription of mental states to oneself. Like Gazzaniga's patients and the head nodders from Chapter 5, a person should often misread her own mind. ISA relies entirely on the verification of both of these predictions: if either turns out to be false, then ISA is absolutely irredeemable.

Moreover, neither of these predictions is compatible with the existence of an inner sense module. For Goldman, there must be people who are fully capable of self-knowledge but who are incapable of mindreading others because they lack the simulative and imaginative capacities. Goldman, with fellow inner sense theorists Nichols and Stitch, looks to autistic people to fulfill this case. And confabulation data is particularly damning for Goldman. Cartesian introspection, at least in the absence of a meddling evil daemon, should never err. The nodders should correctly introspect their beliefs, regardless of whether or not they are nodding.

I will now argue that, with regards to these two predictions, the scientific evidence heavily favors ISA over inner sense theories. ISA predicts that development, genetic anomaly, brain damage, or any other change which inhibits or boosts 3rd person mindreading ability should also equally inhibit or boost 1st person mindreading ability, and vice versa. If the mindreading faculty is a singular module which deals with both one's own and others' minds, then it should improve at mindreading others directly proportionally to its improvement at mindreading oneself, and degradation should be similarly proportionate.[76] As far as normal human development goes, this does seem to be the case.

If mindreading is innate and modular, as I argued in Chapter 6, then the mindreading module should develop the ability to attribute mental states to others concurrently with the ability to recognize one's own mental states. Because human infants consciously draw attention to their own intentions, desires, and so on from an early age, evidence that shows conscious understanding of others' mental states at a similarly early age is required to vindicate this prediction. As Goldman contends, the classic Wimmer and Perner studies showing false-belief understanding at the age of 4 do not cut it. The Baillargeon spontaneous-response tasks, on the other hand, show the ability to attribute propositional attitudes to others at a very young age and hence constitute just such evidence.

As noted in Chapter 6, it is uncontroversial that autistic people have significant trouble with mindreading tasks, and perhaps, in severe cases, lack mindreading capacities altogether. They are thus a prime source for dissociation data. Unfortunately, it is controversial whether autistic people struggle with ascribing mental states to themselves,

[76] Note that this prediction only applies to changes to the mindreading system itself, rather than the perceptual systems that provide its inputs. If one goes deaf, then 3rd person mindreading is likely to suffer to a greater degree than 1st person mindreading, and if one loses to ability to produce inner-speech then the opposite is likely. But the fact that 3rd person and 1st person mindreading use different globally broadcasts perceptions to various degrees does not count against them being produced by the same module. Such, in fact, is the power of the global broadcast system in the first place.

with evidence seemingly supporting such struggle and other evidence seemingly denying it.[77] If people with autism, despite struggling on mindreading tasks, show normal understanding of their own mental states, then ISA is a bad account. If, on the other hand, autistic people struggle with ascriptions of self-belief as well as other-belief, then inner sense theories are called into doubt. Indeed, inner sense theorists Nichols and Stitch (2003) hold that one of the best reasons for upholding the existence of an introspective faculty is that while autistic people have considerable trouble attributing mental states to others, they are able to attribute such states to themselves normally.

Nichols and Stitch rely upon studies that, while generally well-run, only questionably support the conclusions drawn therefrom. For example, Nichols and Stitch take any study that shows autistic individuals successfully reporting their own mental states to be support for their claim, and the vast majority of the studies they reference use adults or children over the age of 11 as subjects. But it is not that case that autistic individuals always fail mindreading tasks. On the contrary, while autistic individuals develop mindreading abilities much later than others, they (whether they are actually engaging in mindreading or simply following heuristics) do show competence in mindreading tasks by the age of 11. Inner sense theory, in predicting a dissociation, does not gain ground by showing that adult autistic subjects can successfully attribute mental states to themselves. Indeed, this is exactly what ISA predicts, for (as is actually concurrently tested in many of the studies) they can also successfully attribute mental states to others.[78]

[77] Carruthers 9, 11

[78] Carruthers 9, 9

Moreover, strengthening this objection is the fact that none of the studies Nichols and Stitch cite actually test autistic subjects' access to their current propositional attitudes. Instead, having been designed for a wide variety of purposes other than proving Nichols and Stitch's hypothesis, they test things like memory formation.[79] There is no reason, under ISA and contemporary memory research, to suspect that memory formation and retrieval involves metarepresentation of the kind carried out by the mindreading system set on oneself. The lack of a mindreading module should not correspond to an inability to create and remember memories. So, all that Nichols and Stitch have proven is that autistic individuals have unimpaired access to kinds of self-knowledge that both ISA and inner sense theories should take them to have.

On the other hand, some studies that actually seem to get at current propositional attitudes suggest that children with autism are poor at self-ascribing mental states. More to the point, a handful of studies show that in general the ability to self-ascribe propositional attitudes comes online, in normal humans as well as autistic humans, at the same age as the ability to ascribe propositional attitudes to others.

For example, in a study by Phillips et alia (1998), children (both autistic and otherwise) were instructed to shoot a toy ray gun at various canisters in hopes of hitting a canister that contained a prize. They guns did not actually fire projectiles of any sort, but instead experimenters surreptitiously (in ways undetectable even by present adults) knocked over predetermined canisters. Before firing the children were asked which canister they intended to shoot, and then after hitting a canister they were asked which canister they had intended to shoot. As ISA predicts, autistic children, especially after

[79] Carruthers 9, 11

winning a prize for shooting a canister they had not originally intended to shoot, were much more likely than normal children to claim to have intended to shoot the canister they in fact hit, but did not in fact intend to shoot. They, just as autistic children performing mindreading tasks have difficulty differentiating between others' intentions and goals, could not differentiate between their own intentions (e.g. to shoot the leftmost canister) and goals (e.g. to shoot the canister with the prize.)

In another experiment, conducted by Williams and Happe (2009), the experimenter asked kids to get a bandaid for her as she pretended to have cut her finger. There were a variety of objects on a close by table, including a clearly marked bandaid box. The kids opened the box, and found not bandaids but crayons inside. When asked what they believed was in the bandaid box, autistic children were much more liable than normal children to respond “crayons”. Just as they are unable to accurately report on others' false beliefs, they were unable to accurately report on their own false beliefs. Indeed, Williams and Happe show that children become able to report their own false beliefs at the same age they would be able to report others' false beliefs. Normally functioning kids claim they expected the box to contain crayons up until around the age of 4, and autistic kids claim the same until the age of 11. Upon reaching those ages, the same ages that verbal reports of mindreading of others come online, the kids were able to acknowledge that they expected the box to contain bandaids.

The neural structures implicated in the 3rd person mindreading system are well-documented and generally agreed upon. There is a mindreading network in the brain – consisting of the medial prefrontal cortex, the posterior cingulate cortex, the temporal pole, the superior temporal sulcus, and the temporo-parietal junction – which lights up on brain imaging scans when one is involved in mindreading tasks. Castelli et alia (2002) note that the same structures light up when a highly functioning autistic individual does mindreading tasks, but to a much lesser degree. Thus, dissociation might be at play if this same neural network is not also implicated in 1st person mindreading.

Brain scanning studies are very difficult work, and most if not all of the studies that have been done regarding self-judgments and other-judgments have clear and significant flaws.[80] The best and most recent brain scan studies are those done by Lombardo et alia (2010). Lombardo's study consisted of asking subjects connected to a brain scanner questions about how likely they were to engage in certain propositional attitudes, such as thinking that keeping a diary is important. Lombardo also asked these subjects how likely the Queen of England was to think that keeping a diary is important. As ISA predicts, in both cases, whether subjects were reporting their own mental states or the mental states of the Queen, the same mindreading network lit up. The scans did show different parts of the network to be active to different degrees in the two trials, and that the network in general was more active in the self condition. Lombardo concludes that this is troublesome for ISA.

But he is wrong. ISA is perfectly compatible with, indeed should predict, the mindreading system being variously and even more involved in 1st person mindreading. After all, ISA claims that one uses more and different sources of perceptual information when self-ascribing mental states than when ascribing them to others, and one's own mental states are likely to carry much more practical weight for that person than the

[80] Carruthers 9, 25

Queen of England's. On the other hand, Lombardo's findings do challenge Goldman's introspection-based simulation theory. Simulation theory actually predicts that 3rd person mindreading should use more brain power than 1st person mindreading, as it involves activating and running the 1st person introspection module plus other imaginative and faculties.

So, based on brain-imaging and autism studies it appears that at the very least there is no strong case to be made that a dissociation between 1st and 3rd person mindreading exists. Indeed, there is growing theoretical evidence that dissociation does not exist. The evidence regarding another of ISA's predictions, that confabulation occurs in interpretation of one's own mind, is much more powerful.

ISA predicts that, given the fact that 3rd person mindreading often involves misinterpretation, so too should 1st person mindreading. Carruthers appeals to an immense amount of evidence to vindicate this second prediction. Not only does the evidence overwhelmingly show subjects messing up when self-ascribing mental states and never being aware of that mess up, but it also shows that such mess ups occur in situations when, given the existence of Cartesian privileged access, the subjects should have easily been able to introspect their actual mental states. Thus, all forms of inner sense theory, and the existence of any sort of mental transparency, are called gravely into doubt.

Recall Gazzaniga once again. The commissurotomy patient in his study was indisputably self-interpreting, and botching the job. Further, because of the transparency assumption, he lacked awareness not only of having botched interpretation, but of having interpreted at all. He simply confabulated his own mental state and went on his merry way.

The only way that inner sense theorists can duck the implications of this confabulation for their work is by doing two things. First, they must stress that, being a commissurotomy patient, the subject was not operating under normal conditions and thus was much more likely to interpret one's own mind than introspect in those conditions. Which is the second thing: that they must not, like Goldman (2006), propose a single introspective faculty. Instead, like Nichols and Stitch, they must propose a two system model of self-knowledge, in which humans possess an introspective module as well as a mindreading module and turn their mindreading module on themselves when, for some reason or another (such as commissurotomy) they cannot introspect. Indeed, the confabulation evidence is so compelling that Goldman himself has come around and is now proposing a two system model.[81] Perhaps the ultimately most compelling general evidence in favor of ISA is that, by and large, over the last 15 years of debate, Carruthers' competitors have substantially altered their theories to more resemble his, rather than vice versa.

But the two system model of inner sense is also quite dubious given numerous confabulation experiments. A number of experiments show that perfectly normally functioning subjects, like the Wells and Petty headnodders, misinterpret their own propositional attitudes even in situations where there is no reason to lack access to introspection (if introspection exists.) For example, the classic Nisbett and Wilson (1997) panty-hose experiment shows that subjects attribute a judgment to themselves that it

[81] See Goldman (2009)

seems clear they did not actually make. In this experiment, subjects, thinking they were involved in a market survey, were presented with four identical items of panty-hose and asked which one they preferred. Despite all of the panty-hose being the same, the subjects showed a very strong bias for the panty-hose on the right hand side (perhaps due to a right-hand attention bias or the English speaker's tendency to scan items left to right). When asked to explain their choices, the subjects immediately offered false reasons, such as softness or color, for picking the rightmost panty-hose.

So, the subjects made a judgment (that the rightmost panty-hose were the best) and then, when asked why they made that judgment, misinterpreted their own reasoning. They did so in the same exact way as they might have misinterpreted an other's reasoning if they had watched someone else pick out panty-hose. Further, unlike in the Gazzaniga experiments, there is no reason for them to have used their mindreading system rather than their introspective module in this situation. No reason, that is, except that introspection does not exist.

My above account of the plausibility of ISA is by no means complete. In his forthcoming book *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Carruthers presents a much more thorough account of ISA, and a great deal more evidence from cognitive science, psychology, and neuroscience both for ISA's truth and against the truth of competing theories. What I have done, I hope, is shown that there exists at least one good theoretical option that maintains that the way by which one believes oneself to be minded is the same as the way by which one believes others to be minded. The remainder of my thesis will be dedicated to examining the implications the truth of such a theory would have for the problem of other minds.

Part IV

Dissolving the Problem of Other Minds

Part I laid out the conceptual and epistemological problems of other minds. Part II laid out my naturalistic methodology. Part III applied this methodology to the conceptual problem of other minds. Now, in Part IV, I explicate my conclusions based on this application. In Chapter 8, with reference to ISA, I solve the conceptual problem of other minds. My solution runs parallel to, and illuminates several key points of, the post-Wittgensteinian solution. But it also differs importantly. Then, in Chapter 9 I show that ISA dissolves the epistemological problem of other minds.

Chapter 8

Solving the Conceptual Problem

I mentioned at the start of Chapter 7 that that chapter, alongside most of the post-Wittgensteinian work, does not (at least explicitly) argue that the concept of mind at work in 1st person and 3rd person ascriptions is univocal. Rather, it argues that the ways in which those concepts are arrived at is univocal; for the post-Wittgensteinians that way is usage in language and for Carruthers that way is the mindreading module. This distinction immediately raises a concern. If I use my mindreading system to attribute a mind both to you and to myself, does that necessarily entail that I attribute the same concept of mind to you as to myself? Thoughtfully applying Carruther's account of interpretive mindreading to Strawson's account of the descriptive metaphysics of personhood ameliorates this worry.

Carruther's evidence shows that the mindreading system attributes the same sorts of mental states (false beliefs, intentions, et cetera) in 1st and 3rd person cases. However, as noted in Chapter 7, there is, given differing amounts and kinds of evidence, a disparity in degree of understanding between my own mind and other minds. There is even a disparity in kind of understanding in that, for example, I interpret my own visual inputs but not yours. But while the sources of informational input to the global broadcast system differ, the processing of that system and its types of outputs do not. Insofar as the mind is active – insofar as we take mindedness (or at least that mindedness which is at stake in the PoOM) to refer to the outputs rather than the inputs of the mind – I attribute all of the same mental states to others as to myself. I do not in general take you nor myself to be mentally capable of any kind of thing of which the other is incapable.

As detailed in Chapter 2, Strawson takes usage in language – the various ways in which we employ the concept of mindedness – to prove that we imagine other minds and our own minds in the same manner. Indeed, our grammar makes it impossible for us to imagine our own minds in any manner different from other minds. Carruthers' work provides theoretical support for the conclusion of this claim, while undermining one of Strawson's beliefs. We use the same language to point out the same concept of mindedness in our own case and others' because the mental process of interpretation is the same in each case. But our concept of mindedness (our own and others') is not robust and simple as Strawson takes it to be, but rather the messy sum of, or one specific instantiation of, our images of our multitudinous interpretations of mindedness.

Strawson assumes that concepts are tidy holistic entities. More importantly, he assumes they are permanent. But, in light of ISA, the notion of 'concept' in Strawson's theory is not one we should take seriously. Concepts are ephemeral packages of acquired information, not holistic objects. Thus, my claim that the concept of 'minded' is univocal is quite different from Strawson's version of the same claim.

Assuming ISA, we come to believe in minds by way of attribution of mental states stemming from the mindreading module. Our working concept of mindedness, for Carruthers, is simply a summation of all of our conscious images of mental states we have attributed. So, because we attribute active mindedness to others by means of the same mechanism by which we attribute it to ourselves, our concept of an other mind is as a matter of fact univocal with our concept of our own mind. In a very Strawsonian move, I take Carruthers' version of global broadcast to entail conceptual identity because it is simply the fact of the matter that we broadcast the concept univocally. The 'grammar' of the concept 'mind', in this case the fact that it is dressed in sensory clothing when globally broadcast, mandates its univocal nature.

Recall Figure A from Chapter 7. Perception writes a certain behavior (whether one's own or that of an other) on the blackboard, which the mindreading system reads off of the blackboard and interprets as a minded behavior. The mindreading system then relays that information to a perceptual system, which writes some version of 'the individual who committed the behavior is minded' on the blackboard. The mindreading system never interacts directly with the mind in question (once again whether one's own, or that of an other), but rather interprets perceptions and feeds attributions of mindedness back into sensory systems which in turn broadcast them as images of mindedness. There is thus no room to substantially differentiate between the 1st and 3rd person concepts. Both concepts of mind are simply processed images of interpretations of the perceived behavior of a subject.

Remembering the discussion at the start of Chapter 7, the post-Wittgensteinians are correct. The way by which one believes oneself to be minded is the same as the way by which one believes others to be minded. And this fact does entail that the concept of mind attributed to oneself is the same as the concept of mind attributed to others. The concept is simply that which announces the results of the method. For Carruthers, the concept 'minded' is the broadcast image of belief in mindedness (whether one's own or an other's). For the post-Wittgensteinian, the concept 'minded' refers to the more primitive concept 'person' which in turn announces that the competent language-user has witnessed someone acting. Again, the witnessed agent may be the post-Wittgensteinian herself or an other person.

Indeed, in this same manner, all of the core post-Wittgensteinian points are vindicated by the science.[82] For example, Avramides' lived position is characterized by engagement and interaction with other agents. Opposed to the lived position is the armchair, from which philosophers posit questions like the PoOM. The armchair mindset is that of the conceptual conservative who, rather than making his philosophy conform with the world, just works off of dubious beliefs and intuitions. Thus, the progressive

[82] This point also works to vindicate Avramides' claim that Strawson's version of the post-Wittgensteinian solution is compatible with Davidson's radical interpretation version. It is by the very act of unconscious interpretation by the mindreading system that we come to understand ourselves and others as minded persons.

naturalist might claim that what is most crucial to the lived position is not interaction, per se, but rather a scientific understanding of the world as a communal place, and of humans as evolved organisms. The lived position, in the naturalist's hands, is merely a call to shed esoteric and dubious concepts and to work instead from concepts given by the world.

Behavior proper and action, then, are just behavior which the mindreading system interprets as the behavior of an agent. Where the mindreading module attributes mindedness, there is behavior proper. The mindreading module attributes mindedness to myself and to others, and this mindedness is a univocal concept. So, I have solved the conceptual problem and established the truth of P3 of the argument from Chapter 2: the concept of mindedness at work in belief in one's own mind is the same concept as the concept of mindedness at work in belief in other minds.

Chapter 9

Dissolving the Epistemological Problem

As noted in Chapter 2, the argument that sets up the epistemological PoOM runs as follows:

P1 One believes that oneself is minded.

P2 One believes that others are minded.

P3 The concept of mindedness at work in P1 and P2 is univocal.

P4 One justifies belief in one's own mind by virtue of direct access to that mind.

P5 One lacks such direct access to other minds.

C One must find another way to justify belief in other minds.

If any one of the premises is false, then the conclusion does not follow, and the problem of other minds does not occur. The post-Wittgensteinian conclusion is that solving the conceptual problem confirms P1, P2, P3 and P4, but undermines P5. According to Strawson and Avramides, by virtue of being a person interacting with other people and using a common ordinary language, one has direct linguo-behavioral access to others' minds as well as one's own mind. Thus, since P5 is false, C does not follow, and the PoOM does not occur.

I disagree with the post-Wittgensteinians on one key point, and thereby make the position much more plausible. P5 is actually true. We do not have special access to other minds. But this is inconsequential because by casting doubt on the transparency assumption, and advancing the indirect sensory-access theory of self-knowledge, Carruthers shows that P4 is false. While the post-Wittgensteinians are correct that 1st and 3rd person access are univocal, they are wrong that it stems from any special sort of direct access. Rather, all belief in mindedness is globally broadcast by sensory systems and indirectly stems from interpretation by the mindreading system. Thus, my response to the PoOM runs as follows:

P1 One believes that oneself is minded.

P2 One believes that others are minded.

P3 The concept of mindedness at work in P1 and P2 is univocal.

P4 One believes in one's own mind by virtue of the mindreading module.

P5 One believes in other minds by virtue of the mindreading module.

C One has no differential reason for justifying belief in one's own mind than one has for justifying belief in other minds.

The classic epistemological PoOM is thus dissolved. Or, more exactly, does not occur in the first place. There is no belief in other minds that requires differential justification than belief in one's own mind. The skeptic is forced to retreat from the problem of other minds to the problem of any minds, and from solipsism to metaphysical nihilism.

Part V

Concluding Remarks

Part I introduced the problem of other minds. Part II introduced my naturalistic methodology. Part III applied this methodology to the problem of other minds. Part IV concluded that the epistemological problem of other minds does not occur. Now, Part V examines the scope and power of my conclusion. Chapter 10 raises and responds to a couple of objections to my view. Finally, Chapter 11 notes that even if my particular conclusions are wrong, my thesis retains importance as an example of progressive naturalistic inquiry.

Chapter 10

A Qualification

To iterate, solving the conceptual problem of other minds is a matter of nailing down the concepts at work in the epistemological problem of other minds, and checking whether or not they make the epistemological problem internally consistent. Obviously, the first of these concepts, and the most difficult to pin down, is simply 'mindedness'. I claim in Chapter 8 that the concept of 'mind' is univocal, but I have not yet provided a clear answer as to what exactly, when one worries about justifying her belief in a mind, one is referring to. The answer that best coheres with a post-Wittgensteinian and/or Carruthersian account is that 'mind' – at least insofar as it is implicated in the PoOM –

refers to the sorts of higher order mental functions, such as propositional attitudes, that are ascribed by the mindreading module. Propositional attitudes, recall, are those mental states that relate persons to propositions. Basically, this sort of answer takes mindedness as that which grants the status of 'subject' to a person. Under this account, the PoOM could be called the problem of other subjects. Or perhaps, in Strawsonian terms, the problem of other persons. Insofar as the epistemological problem of other minds is identical to the problem of other subjects or persons, it is easily dissolvable by reference to the work of Strawson and Carruthers.

But it is conceivable that the PoOM refers to a concept of 'mind' broader than, or perhaps entirely distinct from, the concept of 'subject'. We currently know too little to know the precise scope of psychological phenomena that fall under ISA. That is, our mindreading module seems most at home attributing beliefs, desires, and other psychological states that are intentional or propositional. Thus, if there are psychological states that are undetectable or unprocessable or uninterpretable by the mindreading module, then they will lie outside the scope of the theory. If the PoOM refers, for example, to the question of whether another person feels emotions, then Carruthers' evidence is not nearly as strong as if it refers to propositional attitudes.

Carruthers allows that, given a dearth of scientific evidence, it is possible that visceral 'feelings' (such as disgust or lust) may not be subject to the mindreading system, but rather transparently accessible to the global broadcast of consciousness. However, Carruthers does produce good evidence that emotions, insofar as they might be introspectable at all, can only be intuited in rough, affective, perceptual form. Any sort of emotional or affective nuance, along with any sort of conceptual content, is understood through interpretation by the mindreading module, and thus falls under our purview.[83] Moreover, I suspect that as more science is done, it will become evident that even the rawest of emotions are always processed in some way before reaching consciousness, and that this specific worry will be ameliorated altogether.

A more problematic conception of the PoOM, for my purposes, is that it is identical not with the problem of other subjects, but with the problem of other mere consciousnesses. Perhaps the most vivid illustration of this sort of scope problem comes from the alleged possibility of automata or philosophical zombies. Zombies are beings who behave, look, and function like humans in every way except that they are not conscious. What if the PoOM does not refer to the problem of justifying belief that others have beliefs, intentions, and other attitudes, but rather to the problem of justifying belief that others have qualia; that it is like something to be an other person, in the same way as it is like something to be oneself? David Chalmers (1996) famously terms the explanation of qualitative conscious experience the 'hard problem'. There exist, Chalmers claims, certain qualia (defined by Dennett as “the way things seem to us”[84]) that characterize consciousness and that are indescribable in physical terms. Indeed, many philosophers posit that there is an explanatory gap at play: that consciousness and qualia simply cannot be explained in physical terms. The problem of other consciousnesses is thus as follows: my consciousness may well have access to my mind in the same way as it has access to your mind, but what evidence do I have that you have a conscious awareness that

[83] Carruthers 4, 33

[84] Dennett 381

similarly accesses any minds?

But it is difficult to see why this is a possibility that the progressive naturalist should worry about. Philosophical zombies are a real worry only if all our functionally characterized psychological capacities – affect, emotion, desire, belief, memory, perception, learning, judgment, et cetera – can constitute a coherent psychology in the complete absence of all subjectively felt qualities. My zombie twin would have to be a mammal, and just as good at learning as I am. Yet it is difficult to imagine any known mammal (including me) learning about its environment without the subjectively felt characteristics that accompany rewards, deprivations, and punishments. We know too much about the actual neural mechanisms of learning, especially the dopamine system in the midbrain[85], to take seriously the claim that learning can occur without vividly felt qualities of various sorts.

Furthermore, we might grant that there are some intuitions that seem to give credence to Chalmers' claim – we might even agree that zombies seem to be conceivable to us – but in that case we must consider the possibility that 'qualia' is a dubious concept. It is dubious by psychological role, in that our best reason to uphold it is intuition. And it is plausibly dubious by descent, since it is a vestige of our Cartesian substance dualist ancestry. So, as naturalists, we must bracket the concept of qualia, and move forward in our inquiry without assuming it exists. Furthermore, if our best scientific accounts do not make room for qualia, then we, as Daniel Dennett argued, should hold that "contrary to what seems obvious at first blush, there simply are no qualia at all."[86] If this is right, then

[85] See Panksepp (1998 and 2009)

[86] Dennett 409

the zombie objection has no real traction against my view.

And, lo and behold, ISA, our best scientific account, makes no room for qualia. ISA holds that one must necessarily be conscious to be a self-aware subject. Self-aware subjecthood is predicated on qualitative consciousness. So, establishing that one is a subject is sufficient for establishing that one is conscious. For Carruthers, consciousness is merely a passive byproduct of global broadcast. All and only the contents of global broadcasts are conscious, and consciousness is explained by the fact of their being broadcast. And, for any propositional attitudes to be expressed through behavior, some sensory information has to be globally broadcast. So global broadcast, and with it consciousness, coincides with (and usually predates) the expression of propositional attitudes. Thus, if I attribute propositional attitudes to an other, I must be attributing consciousness to that other. If I show that there is no greater problem with belief in an other's subjecthood than my own, then there is likewise no greater problem with belief in an other consciousness than my own.

Now, there are cases where this does not hold true. As mentioned in Chapter 7, priming experiments show that our perceptual systems affect our actions in ways which are not consciously accessible. Humans thus sometimes behave without being consciously aware of behaving. In this case, my mindreading module might interpret an other as being a subject (and thus conscious) when conscious states do not actually accompany that other's propositional states. But priming experiments show precisely that my mindreading module might just as easily interpret myself as being a conscious subject when conscious states do not accompany my propositional states! In the Bargh, Chen, and Burroughs study, subjects were not consciously aware of their slow gait.[87] In these sorts of cases (and given the huge scope of priming experiments, these sorts of cases are likely constantly occurring) I am a philosophical zombie. I function like a subject, yet am not conscious of that functioning. So there is no greater reason to believe an other is a zombie than to believe that I myself am a zombie.

There only remains one move for my opponent to make: to insist that she is not talking about physical possibility, but logical or conceptual possibility. I have two replies. First, no one has given a non-question-begging account of conceivability that establishes that zombies are possible in any sense at all. Surely we can think we have conceived of some state of affairs when in fact we have not. So, until we have a defensible account of conceivability, we must withhold authority from Chalmers' claim. Second, even if we had a defensible account of conceivability, and even if zombies were conceivable, it is hard to see what bearing that would have on the PoOM. The traditional epistemological problem of other minds is the problem of justifying that those around me have mental states just like my own. It is not merely the problem of justifying that those around me have qualia. This is a point even Chalmers should concede. So, at the very worst, if my view fails to apply to qualia it certainly does not follow that the view fails to resolve a central part of the PoOM.

Moreover, even if this very worst case is true – even if the PoOM is in fact the problem of other consciousnesses, rather than the problem of other subjects, and, contrary to the above argument, self-aware subjecthood is not necessarily predicated on consciousness – a potential naturalistic response is still on offer. Namely, the science

[87] Bargh, Chen, and Burroughs 19

shows that the infamous argument from analogy might not be such a bad response to the PoOM after all. The argument from analogy is the classic response to the PoOM that I can justify belief in other minds by noting that I have a mind and that others are like me in myriad ways and thus probably also have minds. Many take this to be a very bad argument because it involves an overly bold induction from one case to many.

But, with extensive research in evolutionary biology on his side, in “Evolution and the Problem of Other Minds” Elliott Sober (2000) presents an argument from analogy and evolutionary parsimony which just might dispel these concerns. His argument, greatly simplified, runs something like what follows. Humans are evolved animals. Through science and probability theory, we have come to know certain things about evolved species, and particularly the evolution of humans. For one, given phylogenetic parsimony and likelihood analysis, any feature of any particular human that is as complex and integral as mindedness is extremely likely to also be a feature of the vast majority of (if not all) other humans. Therefore, the argument from analogy is not as bad an argument as philosophers make it out to be; it is in fact not an argument from analogy at all but instead a very good inference to the best explanation. While in any given case it is logically possible that an other lacks a mind, we have very strong inductive evidence to believe that she is minded.

So, to summarize this chapter, I must qualify my thesis by noting that there is a conception of the PoOM – as the problem of other mere consciousnesses – that might escape my dissolution. There is reason to believe that it does not, as an idea of unconscious self-aware subjecthood is incoherent on the Carruthersian account. Further, even if a version of the PoOM does escape my analysis, it might be dealt with neatly by Sober's argument from analogy and evolutionary parsimony. Nonetheless, my thesis is strongest if the epistemological PoOM taken to be problematic is the problem of other evolved subjects.

Chapter 11

Towards a New Naturalistic Paradigm

For reasons touched on throughout this thesis, my overarching positive claim – that the epistemological problem of other minds does not occur – is the conclusion of a line of thought that is admittedly not a knockdown argument. Most controversially, my argumentation relies on the uncertain assertions that Carruthers' ISA theory is correct, and that the important PoOM is the problem of other evolved subjects rather than the problem of other merely phenomenal consciousnesses. Despite the scientific evidence I have produced to the contrary, we may have introspective access to our own minds. After all ISA is a cutting-edge scientific theory, and any cutting-edge scientific theory is, simply by virtue of its novelty, likely to be importantly flawed. Likewise, despite my arguments in Chapter 10, the overriding concern regarding other minds might merely be that they comprise, in part, raw phenomenal consciousness that cannot be explained in physical terms.

If either one of these things is true, then my conclusion does not follow. However, with the post-Wittgensteinians, I have no compelling reason to take either of them as true. And, moreso than the post-Wittgensteinians, I have compelling evidence that they are false. At minimum, I take this thesis to make it plausible that knowledge of other minds should not constitute a special problem.

However, whether or not my argument succeeds with respect to the PoOM, I hope that this thesis is successful as a paradigmatic example and extension of Daviesian naturalistic inquiry, as detailed in *Subjects of the World*. More ambitiously, I hope that the way in which I arrive at my conclusion might act as a model for future work in philosophy. As I discussed in Chapter 4, the scientist will probably never be able to satisfy the radically skeptical epistemologist with her account of a human knowing another human's mind. However, I take myself to have provided an example of how the scientist might be able to convince the skeptic that she has as much reason to doubt her own mind as an other human's. And this powerful stroke against solipsism has clear philosophical import.

The shell of my argument runs something like the following:

P1 Philosophers should not assume the existence of dubious concepts, unless said concepts are vindicated by science.

P2 Philosophical debate X assumes concept Y.

P3 Concept Y is dubious for such and such reasons.

P4 The science does not support the existence of concept Y.

C Philosophical debate X should actually be philosophical debate X minus Y.

In particular, I conclude that the PoOM should have the dubious concept that one's own mind is a special case subtracted from it. The main implication of this conclusion for my thesis is that the real epistemological problem about minds pertains to the existence of mindedness itself.

This basic argumentative form might be fruitfully applied to various wrongheaded approaches to issues in philosophy, ranging from externalism about moral reasons to libertarian notions of free will. In my view, good naturalism and insistent interdisciplinarianism do not boil down to an attempt to replace philosophy with science. Rather, they constitute an attempt to make sure philosophers focus on problems that are really worth their time and expertise. And this attempt crucially involves bracketing dubious concepts.

Insofar as the philosopher is asking a particular question (in this case the problem of other evolved subjects), and insofar as a particular scientific account is correct (in this case the indirect sensory-access theory of self-knowledge), the philosopher and the scientist can speak to and not past each other. It is possible that I have represented the PoOM inaccurately, and ISA may well be dead wrong. Nonetheless, there is a certain philosophical problem (namely the PoOM as I represent it) that definitely can be either vindicated (if some sort of modular transparent access is found to exist) or dissolved (if ISA is accurate) by science. Whether or not it does so satisfactorily in this thesis, science can inform philosophy. And this thesis acts as a blueprint illustrating how.

Acknowledgments

First and foremost, I would like to take this opportunity to acknowledge my dad for his encouragement, insightful commentary, and willingness to serve as my foil in Chapter 5. My father's influence on my life has been unmistakably profound. I happily admit to being conceptually imperialistic when it comes to preserving the core Curry values of honesty (intellectual and otherwise), thoughtfulness, and family.

Next, I would like to express my appreciation for my adviser, Paul Davies. His influence – over my entire college career but especially over the past year – has helped me refine and come to peace with my naturalistic worldview. Moreover, even setting aside the central role *Subjects of the World* plays in my argument, this thesis would not be halfway decent without his hard work, unwavering enthusiasm, and engaging conversation.

I would also like to thank Peter Carruthers for graciously providing me with a manuscript of *The Opacity of Mind,* Brian Huss for his constructive advice and encouragement very early on in the project, Tess Curry for her help preparing colloquium and conference presentations, and the rest of my family and friends for their love and support.

References

Aiello, L. and Wheeler, P. 1995. “The expensive tissue hypothesis.” *Current Anthropology*, 36, 199-221.

Avramides, Anita. 2001. *Other Minds*. New York: Routledge.

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

Baars, Bernard. 2010. *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience.* Second Edition. San Diego: Academic Press

Bacon, Francis. 1620. *The New Organon.* Cambridge: Cambridge University Press.

Baillargeon, R., Scott, R. M., & He, Z. 2010. “False-belief understanding in infants.” *Trends in Cognitive Sciences*, 14, 110-118.

Bargh, J. A., Chen, M., & Burrows, L. 1996. “Automaticity of social behavior: Direct effects of trait construct and stereotype priming on action.” *Journal of Personality and Social Psychology*, 71, 230-244.

Baron-Cohen, Simon. 1995. *Mindblindness: An Essay on Autism and Theory of Mind.* Cambridge, MA: MIT Press.

Briñol, P. and Petty, R. 2003. “Overt head movements and persuasion: a self-validation analysis.” *Journal of Personality and Social Psychology*, 84, 1123-1139

Call, J. and Tomasello, M. 2008. “Does the chimpanzee have a theory of mind? 30 years later” *Trends in Cognitive Sciences* Vol.12 No. 5.

Carruthers, Peter. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. London: Oxford University Press.

Castelli, F., Frith, C., Happé, F., and Frith, U. 2002. “Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes.” *Brain*, 125, 1839-1849.

Chalmers, David. 1996. *The Conscious Mind*. London: Oxford University Press.

Chisholm, Roderick. 1964. “Human Freedom and the Self.” The Lindley Lecture, University of Kansas.

Clarke, Desmond. 1992. “Descartes' Philosophy of Science.” *The Cambridge Companion to Descartes*, 267-285. Cambridge: Cambridge University Press.

Darwin, Charles. 1859. *The Origin of Species*. New York: Barnes and Noble Classics.

Davies, Paul Sheldon. 2009. *Subjects of the World: Darwin's Rhetoric and the Study of Agency in Nature*. Chicago: Chicago University Press.

Davidson, Donald. 1997. "Seeing Through Language." *Thought and Language*, Royal Institute of Philosophy Supplement: 42, Cambridge: Cambridge University Press.

Dehaene, Stanislas. 2007. *A Few Steps Toward a Science of Mental Life.* Lecture Transcript. International Mind, Brain, and Education Society and Blackwell Publishing, Inc

Dehaene, S. and Naccache, L. 2001. "Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework." *Cognition*, 79, 1-37.

Dennett, Daniel. 1988. "Quining Qualia" *Consciousness in Modern Science.* London: Oxford University Press 1988.

Descartes, René. 1637. *Discourse on Method*. London: Everyman Paperbacks.

Descartes, René. 1641. *Meditations on First Philosophy.* Cambridge: Cambridge University Press.

Descartes, René. 1644. *Principles of Philosophy.* Cambridge: Cambridge University Press.

Descartes, René. 1649. *Passions of the Soul*. Cambridge: Cambridge University Press.

Descartes, René. 1664. *The Treatise of Man.* Cambridge: Cambridge University Press.

Desimone R. and Duncan, J. 1995. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience.* 18:193-222

de Villiers, Jill. 2007. "The interface of language and Theory of Mind", *Lingua*, Volume 117, Issue 11, *Language Acquisition between Sentence and Discourse*, November 2007, Pages 1858-1878

Gazzaniga, M., and Ledoux, J.. 1978. *The Integrated Mind.* New York: Plenum.

Goldman, Alvin. 2006. *Simulating Minds*. London: Oxford University Press.

Goldman, Alvin. 2009. "Replies to the commentators." *Philosophical Studies*, 144, 477-491.

Gordon, Robert. 1986. “Folk Psychology as Simulation.” *Mind and Language*, 1, 158-71.

Hare, B., Call, J., and Tomasello, M. 2006. “Chimpanzees deceive a human competitor by hiding.” *Cognition*, 101, 495-514.

Hassin, R., Uleman, J., and Bargh, J. eds. 2005. *The New Unconscious*. London: Oxford University Press.

Hrdy, Sarah. 2009. *Mothers and Others: The Evolutionary Origins of Mutual Understanding.* Cambridge MA: The Belknap Press of Harvard University Press.

Leslie, Alan. 1994. “ToMM, ToBy, and Agency: Core architecture and domain specificity.” *Mapping the Mind,* Cambridge University Press.

Leslie, Alan. Friedman, O. and German, T. 2004. “Core mechanisms in “theory of mind””. *Trends in Cognitive Sciences*, 8, 528-533.

Lombardo, M., Chakrabarti, B., Bullmore, E., Wheelwright, S., Sadek, S., Suckling, J., MRC AIMS Consortium, and Baron-Cohen, S. 2010. “Shared neural circuits for mentalizing about the self and others.” *Journal of Cognitive Neuroscience*

Melis, A., Call, J., and Tomasello, M. 2006. Chimpanzees (*Pan troglodytes*) conceal visual and auditory information from others. *Journal of Comparative Psychology*, 120, 154-162.

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

Nijhawan, Romi. 1997. “Visual Decomposition of color through motion extrapolation.” *Nature* 386:66-69

Nisbett, R. and Wilson, T. (1977). “Telling more than we can know.” *Psychological Review*, 84, 231-295.

Onishi, K. and Baillargeon, R. 2005. “Do 15-month-olds understand false beliefs?” *Science*, 308, 255-258.

Onishi, K., Baillargeon, R., and Leslie, A. 2007. “15-month-old infants detect violations in pretend scenarios.” *Acta Psychologica*, 124, 106-128.

Panksepp, Jaak. 1998. *Affective Neuroscience.* New York: Oxford University Press.

Panksepp, J and Northoff, G. 2009. “The Trans-Species Core Self: The Emergence of Active Cultural and Neuro-ecological agents through self-related processing within subcortical-cortical midline networks.” *Consciousness and Cognition* 18. 193-215.

Phillips, W., Baron-Cohen, S., and Rutter, M. 1998. “Understanding intention in normal development and in autism.” *British Journal of Developmental Psychology*, 16, 337-348.

Premack, D. and Woodruff, G. 1978. “Does the chimpanzee have a theory of mind?” *Behavioral Brain Science* 1, 515–52

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

Ryle, Gilbert. 1949. *The Concept of Mind*. Chicago: New University of Chicago Press.

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

Scott, R. M., Baillargeon, R., Song, H., & Leslie, A. M. (in press). “Attributing false beliefs about non-obvious properties at 18 months.” *Cognitive Psychology.*

Searle, John. 1983. *Intentionality.* Cambridge: Cambridge University Press.

Senju, A., Southgate, V., White, S., and Frith, U. 2009. “Mindblind eyes: An absence of spontaneous theory of mind in asperger syndrome.” *Science*, 325 (5942), 883-885.

Shanahan, Murray. 2006. “A cognitive architecture that combines internal simulation with a global workspace” *Consciousness and Cognition* 15 (2006) 433–44

Sober, Elliott. 2000. “Evolution and the Problem of Other Minds.” *The Journal of Philosophy*, Vol. 97, No. 7. 365-386

Song, H. and Baillargeon, R. 2008. “Infants’ reasoning about others’ false perceptions.” *Developmental Psychology*, 44, 1789-1795.

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

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

Strawson, Galen. 1994. *Mental Reality*. New York: MIT Press.

Strawson, P.F. 1959. *Individuals*: *An Essay in Descriptive Metaphysics*. Bristol: J.W. Arrowsmith Ltd.

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

Tomasello, Michael. 2008. *Origins of Human Communication*. MIT Press.

Varley, R. 1998. “Aphasic language, aphasic thought.” In P. Carruthers and J. Boucher (eds.), *Language and Thought*, Cambridge University Press.

Varley, R., Siegal, M., and Want, S. 2001. “Severe impairment in grammar does not preclude theory of mind.” *Neurocase*, 7, 489-493.

Wallace, David Foster. *This is Water*. 2009. New York: Little, Brown and Company.

Wegner, Daniel. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Wells, G. and Petty, R. 1980. “The effects of overt head movements on persuasion.” *Basic and Applied Social Psychology*, 1, 219-230.

Williams, D. and Happé, F. 2009. ““What did I say?” versus “What did I think?”: Attributing false beliefs to self amongst children with and without autism.” *Journal of Autism and Developmental Disorders*, 39, 865-873.

Williams, D. and Happé, F. 2010. “Representing intentions in self and other: Studies of autism and typical development.” *Developmental Science*, 13, 307-319.

Wilson, Timothy. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious.* Cambridge, MA: Harvard 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 (1): 103–128.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations.* New York: Prentice Hall.