

THE REPRESENTATIONAL CHARACTER OF IMAGINATION

by

PETER LANGLAND-HASSAN

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2009

© 2009

PETER LANGLAND-HASSAN

All Rights Reserved

This manuscript has been read and accepted by the
Graduate Faculty in Philosophy in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

_____	_____
Date	David M. Rosenthal Chair of Examining Committee

_____	_____
Date	Iakovos Vasiliou Executive Officer

Jonathan E. Adler

Michael Levin

David M. Rosenthal

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

The Representational Character of Imagination

by

Peter Langland-Hassan

Adviser: Professor Jonathan Adler

Two dogmas shape most theorizing on sensory imagination (thought involving imagery) and propositional imagination (imagining *that* thus and such). The first is that imaginers have privileged access to what they are imagining; the second is that imagining involves cognitive mechanisms over and above those underlying belief. I challenge both assumptions, arguing that one can easily be wrong about what one is sensorily imagining, and that propositional imagining requires only ordinary beliefs and desires. The former claim is supported through a distinction between the representational (or ‘intentional’) content of an imaginative experience and the matter of whether the “success” conditions given by that content are satisfied. The latter is advanced on grounds of parsimony, as more baroque hypotheses are shown not to be borne out by the data. In addition, a novel theory of the cognitive mechanisms underlying the sense of agency had over one’s own imaginings is developed, through an analysis of cases (in schizophrenia) when the phenomenology of thought-agency is abnormal.

The cumulative effect is to replace the view of imagination as a *sui generis*, “off-line” mental phenomenon with one that sees it as an assertoric faculty aimed at representing past experiences and future possibilities.

For Antoinette and Avery – my centers of gravity, and levity

ACKNOWLEDGEMENTS

It's customary to incur lots of debts when writing a dissertation, and I offer no exception. My most obvious debt is to my adviser Jonathan Adler, who picked up the project midway through and, without a moment's arm-twisting, allowed me to continue on with my overly ambitious plan. I learned as much from his personal warmth, graciousness, and open-mindedness as I did from his astute critiques of my work. Special thanks are due also to Michael Levin, whose line-by-line emendations made me a better writer, as well as philosopher. Galen Strawson, David Rosenthal, and Jesse Prinz also made their way through complete drafts and provided valuable feedback, both before and during the defense.

I offer my deepest thanks to my wife Antoinette, who encouraged me to pursue a career in philosophy, and ever since has steadfastly absorbed the sacrifices of marrying—and raising a child with—an underemployed philosopher. To our two-year-old son Avery: thank you for your strict adherence to a predictable nap schedule, during which much of this was written; and thank you for keeping a smile on my face so much of the rest of the time (and forgive me if, from time to time, I muttered under my breath about “representational content” as you played in the Harmony Playground sandbox).

Finally, I thank my Mother and Father, who have always made me feel like I know what I'm doing. It has proved a useful illusion.

TABLE OF CONTENTS

	Page:
Introduction	1
Chapter 1: Failures of Imagination.....	11
Chapter 2: The Contents of Imagination.....	60
Chapter 3: The Appearance of Possibilities.....	97
Chapter 4: Propositional Imagination: Imagining as Believing.....	143
Chapter 5: The Phenomenology of Imaginative Agency.....	195
Summary and Speculation	246
References	253

INTRODUCTION

This project began as an investigation into a few puzzles—puzzles all loosely connected to the topic of imagination. I did not expect any very general insights into the nature of imagination to emerge—nor did they seem possible. There were too many different things called ‘imagination’, by too many people.¹ Philosophers tend to treat ‘imagination’ like a government tax refund—use it for that special something, without much need for justification or rationalization. Go ahead and stipulate.

Despite the many referents of ‘imagination’ in the philosophical literature, there seemed to me at least three core phenomena doing business under the name ‘imagination’. First, there was mental imagery. One tradition in philosophy and psychology has it that imagining is a matter of forming mental images, typically keyed to some sense modality or other (recently the term ‘sensory imagination’ has been used to pick out this phenomenon). Visualization and hearing a song “in your head” are typical examples.

Then there is what philosophers have called “propositional imagination”. Propositional imagination is closely tied to the notions of pretense, supposition, fantasy, and hypothetical reasoning more generally. To engage in such activities one need not form mental images (though one can); one need only direct one’s mind toward what is not actual, toward the merely possible. So the phenomenon of propositional imagination is, in essence, quite distinct from sensory imagination. Nor, might I add, can sensory imagination straightforwardly be seen as a subspecies of propositional imagination. After all,

¹ For an exceedingly thorough account of the many philosophical uses of ‘imagination’, see Brann (1991).

autobiographical memory involves lots of mental imagery, and remembering does not in any obvious way require pretense, supposition, or fantasy.

Thirdly, there are (what I will call) the agency phenomena. Appealing to imagination's relation to the will has traditionally been an important way of distinguishing imagination (both sensory and propositional) from other thought processes, such as perception and sensation. Someone interested in imagination will thus naturally be interested in the nature of thought agency. Questions in this area include: what, phenomenologically, does the *feeling* of being the agent of one's thoughts consist in? What implications does such agency have for the representational content of our thoughts? Does willful control over our imaginings entail that we always know what we are imagining? Can we imagine things we believe ourselves not to be imagining? These questions arise in the consideration of both sensory and propositional imagination, yet also extend beyond the purview of each, as they concern the nature of thought-agency in general. Thus it seems right to consider it a topic in itself—one closely connected to the issue of imagination (broadly speaking), but perhaps most fruitfully explored independently.

So, I set out to explore imagery, possibility, and thought-agency separately, but in the context of a broad literature on imagination. And yet, like characters in a romantic comedy, they seemed to keep running into each other in unexpected places, revealing unsuspected connections. This introduction is a brief account of how and where they rendezvous in the chapters to come, and an indication how the resultant relationships enable more generalized approach to "imagination" than I originally thought possible.

My initial interest was not in imagination as such, but rather in mental imagery's role in reasoning about the mind/brain relation. It always struck me that mental imagery played a

key (and not entirely innocent) role in generating the conundrum of the mind's relation to body and brain. As others have remarked², there is a tendency to use very different images when imagining experiences “from the inside” than when imagining the token brain states to which they may be identical. Might this represent a way of tracing the intuitive mind/brain “gap” to the way that many are inclined to *think of* the relation? The whole business of imagining things “from the inside”, as opposed to some other way (“from the *outside*”?), seemed to me dubious. But then, maybe we use different images to imagine experiences and squishy brain states for the simple reason that they really are quite different and distinct.

It is hard to see how one might gain real traction on the question, so long as it is tacitly accepted that people can generally imagine what they wish to imagine, and usually know what they are imagining when they imagine it. Of course, most will admit that imagination *can* lead us astray in certain instances; but on what basis could the materialist—without begging the question—say that the mind/brain case was one such instance? The dialectic seems doomed to end in a staring contest: one side insists that imagination sometimes misleads, the other that there is no reason to think it misleads in *this* instance.³

At some point it occurred to me that there is a range of more pedestrian—and typically overlooked—cases where we freely admit that a person was wrong about what she was imagining (I am still discussing only *sensory* imagination). I might deny that I am imagining a fruit when I imagine a tomato, if I don't know that tomatoes are fruits; nevertheless, if tomatoes really *are* fruits, it seems right to say that I am imagining a fruit

² Nagel (1974, fn. 11), with his perceptual imagination/sympathetic imagination distinction, is the *locus classicus* of this sort of view. Hill (1997) more recently expands on Nagel's line of argument. Both developments have been criticized by a number of authors including Stoljar (2006, pp. 191-197), Chalmers (1999), and Papineau (2007).

³ See, e.g., the exchange between Hill & McGlaughlin (1999) and Chalmers (1999).

when I imagine a tomato, regardless of whether I say that I am. This leads to the following puzzle: on the one hand it seems that the subject matter of our sensory imaginings is “up to us”—we determine through our intentions what it is we are imagining, in apparent contrast to instances of perception. On the other hand, it also seems we can (rather easily) be wrong about what we are sensorily imagining, just as we can be wrong about what we are seeing or hearing.

Thinking about this puzzle, which lies at the heart of Chapters One and Two, raised so many questions for me that, by the time I got back to my original mind/brain query (which I address at the end of Chapter Three), it was somewhat as an afterthought. Of course, it’s not necessarily a bad thing to unwittingly back one’s way into a position on a contentious issue. It may even lend the view—right or wrong—an air of objectivity.

My approach to the puzzle is to argue that we can indeed be (radically) wrong about what we are sensorily imagining, and to try to show how this is compatible with the sense we have of being the authors or agents of our own imaginings. Thus, the issue of agency is already front and center in this first chapter that was initially just to be about imagery; the question is, what epistemological privileges (if any) does being the agent of an imagining allow one? It is not until the end of Chapter Three that my answer is completely spelled out. A large part of the answer, it turns out, involves explaining how we represent possibilities. This introduces the third character in our romantic comedy. For if we use imagery to represent possibilities, and if some things we represent as possible are *not* possible, then one can begin to see how our sensory imaginings can fail to hit their mark—can *misrepresent*—even if they are generated through our own agency, and out of our own intentions.

An important conclusion of Chapter One is that sensory imaginings—in the visual instances on which I focus—represent the way things looked, or would look. This is in contrast to the many current views that analyze imagination as having essentially the same representational contents as perceptual states (though run “off-line”), as well as to the view that imaginative experiences are “fitless” (i.e., lacking a direction of fit) and only “as if” they have correctness conditions (Searle, 1983). On my view, there is nothing “off-line” or “fitless” about the representations featured in imaginings. Chapter Two goes into explicit detail concerning the representational (or “intentional”) contents of such states. I argue that sensory imaginings have a complex content, involving an existentially quantified content (playing the role of the “image”) conjoined to a “causal-indexical” content. The account of these contents makes sense of a range of both intuitive and empirical data concerning the role of imagination in reasoning, memory, and action-planning. I also compare my account to some prominent contenders and respond to some potential objections.

Chapter Three tackles the thorny question of whether, and to what degree, imagination is a “guide” to possibility, moving me back in the direction of my original mind/brain query. I look specifically at Kripke’s (1980) famous arguments for *a posteriori* necessities, as these give rise to a range of perplexing questions concerning the relation between imaginability, possibility, and modal intuition. I argue that his means for explaining away various mistaken modal intuitions (by appeal to “qualitatively identical epistemic situations”), together with his remarks about what can and cannot be imagined, suggest that sensory imagination plays a key (if inexplicit) role in his modal epistemology—a role that fits well with the conception of sensory imagination I develop in Chapters One and Two. Once I have put Kripke’s strategy for explaining away mistaken modal intuitions into my

own terms, I show how his arguments against mind-brain identity theories can be defused in the same manner in which he explains other mistaken intuitions away (a feat he famously argues cannot be done). Along the way I argue for limits on the kinds of modal determinations acts of sensory imagination can reliably inspire.

The connection between imagery and possibility is obvious in Chapter Three, but the issue of agency is lurking there as well. An overlooked question in modal epistemology is whether imaginings *just are* our beliefs and commitments about what is possible, or whether they are some other kind of state that itself rides free of such commitments. Most philosophers assume the latter (“riding free”) view, whether or not they agree that such states act as reliable guides to (or evidence for) possibility. It is typically thought that imaginings cannot just *be* beliefs since imaginings are under control of the will, while beliefs (and assertoric states more generally) are not.

But if there is a way to understand imaginings as both subject to the will *and* as constituting (rather than simply guiding) our beliefs concerning possibilities, the picture changes in important ways. I show how such a view of imagination may be correct in Chapters One and Two, and draw out the implications of this view for our determinations of possibility in Chapter Three. A large part of understanding imagination’s relation to possibility requires first getting clear on what does and does not follow from the will’s involvement in guiding imagination.

Chapter Four begins what I originally conceived of as a completely independent discussion of propositional imagination. To propositionally imagine something is to imagine *that* x , where x is some proposition (e.g., ‘the sky is falling’). This is opposed to non-propositional uses of ‘imagine’ where someone simply imagines X (where X is some object),

and which typically mark the presence of imagery. Propositional imagination includes cases one might also describe as supposition and hypothetical reasoning; there need be no imagery used when someone propositionally imagines something. For example, one can propositionally imagine *that it is hot in Antarctica* without forming an image.⁴

In opposition to the current wave of theories that hold that propositional imagination involves a distinctive cognitive attitude (and correspondingly distinct elements of cognitive architecture), I argue that propositional imagination only involves belief and desire, and that the data motivating other views can be explained without positing elements of cognitive architecture beyond those underlying belief and desire. Most of the relevant data arises out of the human ability to engage in pretense. Nothing in such behavior, I argue, requires positing a distinctive cognitive attitude of ‘imagination’. A key claim of my view is that neither pretending that *p* nor imagining that *p* require entertaining the proposition that *p*. Thus, in instances where one pretends something that one disbelieves, there is no need for the pretended proposition to be “quarantined” (via a distinctive cognitive attitude) from one’s conflicting beliefs. Again the issue of agency arises: how can (propositionally) imagining be a matter of simply believing various things, if what we propositionally imagine is “up to us”? Again I argue the conflict is only apparent. Imagining that *p* consists in (actively!) forming beliefs about what would be the case if *p*, based upon retrieved beliefs concerning relevant conditionals and generalizations.⁵ The “freedom” of imagination consists in our ability to choose the subject matter of this reasoning, not the specific beliefs we thereby generate.

⁴ Whether an act of supposition or hypothetical reasoning must be sufficiently “rich” before it counts as an act of propositional imagination is a matter I will leave to the side—nothing of substance in my discussion hangs on it.

⁵ This is oversimplifying a little. I argue in Chapter Four that imagining that *p* also requires not believing that *q*, where *q* is part of the same imaginative project (and where *p* may or may not = *q*). More on the nature of *q* in Chapter Four.

Since many cases of propositional imagination involve mental imagery (albeit non-essentially), it is natural to desire an account of how representational contents distinctive of both sensory and (non-sensory) propositional imagination can combine during a single imaginative project. At the end of Chapter Four, I describe one such project at the level of representational content.

Amazingly, it was only once my work on propositional imagination was well underway that I realized I was, in essence, making many of the same claims I'd made concerning sensory imagination. Most importantly, my reduction of propositional imagination to inference and belief cohered remarkably (suspiciously?) well with the assertoric account of sensory imagination I'd developed in earlier chapters. But whereas the account of sensory imagination was born out of the puzzle of Chapter One, the view of propositional imagination is motivated primarily by empirical work on pretense, and by considerations of simplicity and parsimony as concern theories of cognitive architecture. If the main argument of Chapter Four is cogent, the data motivating most theories of propositional imagination simply do not warrant positing cognitive mechanisms that extend beyond those underlying belief and desire.

Thus, via two very different paths—and in an attempt to treat two distinct topics *as* distinct—I arrive at more or less the same place: an assertoric theory of (sensory *and* propositional) imagination. According to me, both sensory imagination and propositional imagination are (with a few qualifications⁶) activities that consist in tokening beliefs and drawing inferences. Since beliefs and inferences are phenomena we have a comparatively

⁶ I do not ultimately *identify* the cognitive “attitude” at work in sensory imagination with belief—but I come very close. I maintain a difference mainly in order to respect the fact that there appear to be separate cognitive mechanisms underlying both “attitudes”. See the discussion at the end of Chapter Two.

good theoretical grasp on, this “reduction” is genuinely explanatory. Far from the traditional views of imagination that see it as a *sui generis* “offline” phenomenon that is disconnected from our actual beliefs and commitments, I argue that imagining—both sensory and propositional—is *constituted by* our actual commitments concerning what would be likely in various instances.

Much of the warrant for positing this reduction lies in its ability to explain a variety of relationships among the three core phenomena of imagery, possibility, and thought-agency. These relationships are further explored the final Chapter Five, where I examine the disorder of thought insertion in schizophrenia. Patients experiencing inserted thoughts claim that other people’s thoughts have been “inserted” into their own minds. The goal of the chapter is to understand, in the context of current cognitive theories of agency, how it is that one could come to believe that one’s own thoughts are someone else’s. This question constitutes what I call “the puzzle of extraneity.” The working assumption of many theorists doing research in this area is that the subjective senses of thought *ownership* and of thought *agency* are dissociable, and that they come apart in patients suffering from thought insertion. The controversial questions are what, exactly, the subjective sense of “thought agency” consists in, and what cognitive mechanisms are responsible for generating it. I try to shed light on these questions in developing a novel approach to the puzzle—one that builds on and amends the earlier theories of Frith (1992), Jeannerod and Pacherie (2004), and Proust (2006), among others.

While this chapter obviously focuses on the issue of thought-agency (and its phenomenology), an important connection is made to the earlier chapters on imagery. Inner speech, as it involves auditory verbal imagery, is a form of sensory imagination. I argue that

both inserted thoughts and auditory verbal hallucinations result from a disorder of inner speech—that they are, at bottom, two ways of reporting the same phenomenon. Importantly, the phenomenological sense of thought-agency does not float free of the imagery and representational contents of the states involved in imagination, but rather is *constituted by* certain ways in which the imagistic portions of the states themselves are processed (here I appeal to an extensive empirical literature on the cognitive mechanisms that underlie the sense of agency in other perceptual and sensory modalities). Again, two of the three core phenomena that initially seemed distinct end up bearing an extremely close—even constitutive—relation. Showing how the sense of agency we have over our own inner speech is generated—and how it can become distorted—goes a long way toward explaining what it is to feel in control of our sensory imaginings in other modalities as well.

So much for coming attractions. Inevitably, many interesting issues often discussed under the heading of ‘imagination’ will be left to the side. I will not have much to say about the role of imagination in responses to artworks (though Chapter Four, on pretense, should make it clear what I *would* say concerning various puzzles in aesthetics). Nor do I wade very far into the dark waters of the traditional “imagery debate” in philosophy and psychology (in Chapter Two, I propose approaching this debate from a different direction). Finally, I have to leave for another occasion the phenomenon of “imaginative resistance” (T. Gendler, 2000).

CHAPTER 1:

Failures of Imagination

1. Introduction

I imagine things.

How?

Well, I pick what I want to imagine and, *god willing*, I imagine it. I imagine things that no longer exist; and, I imagine things that have *never* existed.

Yet I can also be wrong about what I am imagining (or “visualizing”, I will use the terms interchangeably⁷). I may think that I am imagining your podiatrist in a pirate costume, while, in reality, I am imagining someone else. It may be your gardener—the one with the unusual moustache—who I am imagining; for it is he who showed me ankle exercises last week, falsely claiming to be your podiatrist.

Is that right? Can I really be wrong about what I am imagining? Certainly I can be wrong about what I am *seeing*. I may think that I am seeing Mars when really it is Venus. In similar ways, I can be wrong about what I know, what I am remembering, or what I am hearing. Yet, when I go wrong in each of these cases, we know at least roughly what it

⁷ Unless otherwise noted, I use ‘imagination’ in Chapters 1-3 to refer to all thought involving visual imagery. What I mean by ‘imagination’ in these cases is what others have called *sensory imagination*, limited to its visual instances (Byrne, 2007; T. Gendler, 2005; Martin, 2002; Noordhof, 2002; Peacocke, 1985; Thomas, 2007). Martin offers the following characterization: “By sensory imagining I have in mind those distinctive episodes of imagining or imaging which correspond to our use of the distinct senses: so we talk of visualizing corresponding to seeing, or listening in one’s head parallel to audition, and so on” (2002, p. 403).

Another way to mark this distinction is to say that I am here interested in *non-propositional* uses of ‘imagine’, such as when one imagines a horse, or imagines the Eiffel Tower. This is opposed to its propositional uses (e.g. imagining *that* the sun is made of butter, or *that* it is raining on the moon), which are more closely linked to the notion of pretense. In Chapter 4 I explicitly discuss propositional imagination and provide my own account of its relation to sensory imagination.

would have taken for me to be *right*. If I can indeed be wrong about what I am imagining, what must be the case for me to be imagining what I believe myself to be imagining?

The peculiarity of that question aside, there are strong intuitions that pull against the claim that we can be wrong about what we are imagining. Many have found it absurd to question whether a person is really imagining what he thinks he is imagining (McGinn, 2004, pp. 30-32; Peacocke, 1985, p. 20; Wittgenstein, 1969).⁸ It is commonly held that imagining is a willful activity, connected to our intentions in a special way, and that therefore what we imagine must be “up to us.”⁹ So then, could we be wrong about what we are imagining if we ourselves determine, through our very intentions, what is imagined? From this perspective, our epistemic access to what we are imagining seems more like our epistemic access to what we believe or desire; only in extraordinary circumstances will one be wrong about such things.

This leaves us with a puzzle: on the one hand, it is often natural to say that a person was wrong about what he was imagining. I thought I was imagining your podiatrist in a pirate costume, but I was not—indeed, I admit as much on later learning of your gardener’s deceptive ways. Moreover, there seems nothing extraordinary about such cases; they involve relatively routine confusions. Yet, on the other hand, what we are imagining seems to be “up to us” in some crucial respect. So how could we so easily be wrong about what we are imagining?

⁸ There are some complexities in Peacocke’s treatment of this issue that require further discussion—see section two of chapter two.

⁹ Cf. Fodor (1975, p. 191): “What makes my stick figure an image of a tiger is not that it looks much like one...but rather that it’s *my* image, so I’m the one who gets to say what it’s an image of.” More recently, Colin McGinn has called it a “familiar point” that “I cannot misidentify the object of my imagining” (2004, p. 31). See also Noordhof’s (2002, pp. 429-434) “Straight Forward View”, which identifies what is imagined with what an imager supposes is imagined.

At first blush this may seem like a kind of pseudo-puzzle, one that can easily be cleared up by appeal to a familiar distinction or two, along with the introduction of a little terminology. I hope to make clear, however, that the puzzle runs quite deep. I will defend what is likely the more controversial of the claims—that we can easily be wrong about what we are imagining—while explaining the other away. I will argue that we can be wrong both when we are imagining things from our (perceptual) pasts and when we are imagining novel scenarios. And I will argue that this is so even taking into account the fact that an imaginer's intentions play a role in determining what is imagined. It is worth noting that my account of how we can be wrong will not involve a simple appeal to our potential ignorance of the causal source of our images—a possibility granted by Peacocke (1985). Nor does it amount to a kind of propositional confusion, discussed by Stoljar (2006, p. 75-77) and others, where one is mistaken about the proposition one is actually entertaining.¹⁰ My discussion will hinge instead on defending a distinction between the representational content of one's imaginative experience and the question of what (if anything) one is *in fact* imagining.

I use 'representational content' equivalently with 'intentional content'. The representational (intentional) content of a thought or experience can be compared with the contents of a newspaper (Siegel, 2005)—it is what the mental state “says” or represents as being true about the world. A belief can have the representational content *that there is a full moon* even if there is not a full moon, just as a newspaper can say things that are false. Most philosophers accept that perceptual states have representational content in a similar way—they can represent the world as being ways it is not. For instance, if (as in the Muller-Lyer

¹⁰ This conception of what it is to be wrong about what one is imagining—which often arises in the context of Kripke's (1980) arguments for *a posteriori* necessities—is treated in detail in Chapter 3. Note that it is most naturally viewed as an account of how we can be mistaken about what we are *propositionally* imagining—not what we are sensorily imagining.

illusion) someone looks at two lines of equal length, but top one looks to be longer than the other due to the different orientation of arrows at its end, the representational content of the visual state is (roughly speaking) *that the top line is longer*. The perceptual state has this (non-veridical) content even though the experience is *caused by* lines of equivalent length. I take these to be fairly customary claims about the nature of representational that do not by themselves weigh one way or the other in favor of internalist or externalist theories of how such contents are determined.¹¹ Moreover, the claims I will make about the contents of imaginative experiences should go through regardless of whether one accepts internalism or externalism about the determination conditions for such contents.

A last objection worth addressing at the outset is that the problem, as I have set it up, misses the point that sensory imagination does not represent particulars—rather, one might object, it represents unspecified things of this or that kind (*a* squirrel, or *a* red ball, not some particular squirrel or red ball). If that were right then it would weigh against the claim that I was wrong about what I was imagining when I tried to imagine your podiatrist. For I simply imagined *a* man in a pirate costume, and was correct so far as that went.

The problem with this objection is that it already presupposes a sophisticated theory of sensory imagination, one no doubt driven by the intuitive datum that a single visual image can serve multiple imaginative projects. I aim to start at a more neutral place, one which accepts—in line with commonsense—that we can in fact imagine particulars. For we normally speak of imagining particulars such as Bill Clinton or The Empire State Building,

¹¹ All major externalist theories of content allow for the possibility of perceptual misrepresentation in one way or another—usually by appeal to either optimal conditions, or to a teleological evolutionary account. For instance, on an evolutionary account, if the kind of visual state that in one's species' evolutionary history was adaptive for discriminating the color red is somehow triggered by a green object, one's visual state nevertheless represents the object as red. On an optimal conditions account, a mental state that is reliably caused by allows one to successfully engage with a certain kind of object or property

not merely a Bill-Clinton-looking person, or a building with an Empire-State-Building-esque appearance. While it remains to be explained how a single mental image can play a role in representing multiple scenarios, we should seek an explanation that respects the equally strong, if not stronger, intuition that we imagine particulars.

The argument of this chapter will proceed as follows. In Section Two, I situate imagination among other *de re* mental acts—such as seeing, remembering, and hearing—by drawing a number of distinctions and introducing some terminology. The most important distinction drawn here is that between an *imaginative experience* (defined by its representational content), and the question of *what is in fact imagined* (where ‘imagine’ functions as a “success verb”). Section Three is devoted to saying why some other common ways of understanding the representational content of imagination are deficient. Part of my argument here involves an appeal to empirical work on the role of imagery in memory and action-planning that will be of relevance throughout the dissertation. Section Four directly confronts the main intuition that clashes with my approach to the puzzle—this is the intuition that what we imagine is a matter of our own stipulation, or, at any rate, that we cannot at all easily be wrong about what we are imagining.

Having shown why the intuition that what we imagine is “up to us” is ill-founded, I conclude in Section Five that imaginative experiences represent how the world looked, or would look, in certain situations, and that we do not have privileged access to the facts concerning whether the world *really* looked, or would look, those ways—hence our liability to error about what we are imagining. This sets the stage for a more detailed account of how these contents are structured, which follows in Chapter Two.

1.1 Is remembering a form of imagination?

Before getting to the main argument of this chapter, I want to address a potential worry concerning how I have set things up. I have defined imagination as visualization or, equivalently, as all visual-imagery-involving thought. So understood, imagination will naturally include cases that could be described as experiential *memories*, as well as cases that involve pure fantasy. One might object that, while we can easily be wrong about what we are remembering (e.g., because our apparent memories are not caused in the right way), no such errors threaten us when we are merely fantasizing. If that is right, then including the two kinds of thought in one class called ‘imagination’ will be misleading; for the chapter’s main claim that we can be wrong about what we are imagining will only apply to a subset of imaginative states.

Be assured that I will indeed grant a difference between imagistic memory and imagistic fantasy, by treating memory cases and fantasy cases somewhat differently. I nowhere assume that, just because something is true of the memory cases it is also true of pure fantasy. Nevertheless, there are factors that weigh against treating the two kinds of cases very differently.

First, on an intuitive, pre-theoretical level, it is far from clear where the line falls between imagery-involving memory and imagery-involving fantasy. If I am remembering my toaster as it looked sitting on the table this morning, and then imagine it slowly rising into the air, do I, at the precise moment it lifts off from the imagined countertop, shift from memory to an importantly different kind of mental state? Or what if I simply go from imagining my toaster on the countertop to imagining some other toaster of the same make on an exactly similar countertop? At a psychological level, it seems that nothing has changed in

these cases that would call for a theoretical firewall between the various mental states involved, even if *some* change has occurred that should be recognized.

Indeed, in many cases it will be difficult to decide whether one is remembering, fantasizing, or doing a little of both simultaneously. This difficulty is often overlooked, as there is a (often tacit) tendency to view *all* visualization as a kind of pretense (perhaps a pretending to see thus and such). But, absent independent argument, this view should be resisted. I certainly do not need to pretend anything when remembering things *without* imagery; there is no obvious reason to think that the presence of an image in memory changes that. Part of the confusion here may be that there is another use of ‘imagine’—its propositional use, when one imagines *that* such and such—that *is* invariably linked to the notion of pretense. But the topic I am discussing in this chapter is *sensory imagination*, which is typically denoted by non-propositional uses of ‘imagine’ (imagining the Eiffel Tower; imagining Bill Clinton, etc.), and which is normally defined as imagery-involving.¹² Propositional imagination will receive its own extended discussion in Chapter Four. When the distinction between propositional and sensory imagination is kept straight, it is often quite difficult to know where to draw the line between sensory imagination that is memory and that which is fantasy. Indeed, there are likely to be shades of gray here, and we should prefer an account that respects this feature of imagination.

A second, independent, argument for thinking that imagery-involving memory and imagery-involving fantasy are importantly linked is that this connection is supported by a wealth of emerging research in cognitive science and neurobiology. In a recent review

¹² Cf. Velleman (2000, p. 250, fn. 11): “To imagine *the moon’s being made of green cheese* may simply be to entertain a thought in the form of a mental image, an image of a green-cheesy moon. But to imagine that *the moon is made of green cheese* is to engage in a mental fiction, which may or may not involve imagery.” See Walton (1990, pp. 19-21) for a similar distinction.

article, Schacter *et al.* (2007) note that “brain regions that have traditionally been associated with memory appear to be similarly engaged when people imagine future experiences.” They believe such observations “may have far-reaching implications for conceptions of memory and its functions,” and that “memory can be productively re-conceptualized in light of this idea” (p. 657).

The evidence for these claims comes from two main sources: patients with memory deficits who are revealed to have similar deficits in imagining the future, and brain-imaging studies that reveal activity in a shared “highly specific core brain system” (p. 659) both when subjects are engaged in episodic memory¹³ and when they are imagining future events. An example of the first sort comes from Hassabis *et al.* (2007), who found that four out of five amnesiac patients with bilateral hippocampal damage were severely impaired in their ability to generate (and describe) rich imaginary scenarios. Also, older adults have been found to be impaired, relative to college students, both in the level of detail contained in episodic memories and in the richness of imagined future events (Addis, Wong, & Schacter, 2008). Buckner and Carroll (2007) provide a useful synthesis of the neuro-imaging studies that weigh in favor of shared neural mechanisms underlying episodic memory and imagining the future.

Just as I combine image-driven memory and image-driven fantasy into a single class called ‘imagination’, so too does Schacter describe a single faculty of “episodic thought” which can occur in “two temporal directions”, one “retrospective”, the other “prospective” (p. 658). Thus, while my way of cutting up the class of imaginative experiences may run

¹³ Psychologists generally use ‘episodic memory’ to refer to autobiographical memories keyed to a specific time and place—what we might think of as memories of personal experiences (Schacter *et al.*, 2007, p. 657). This faculty is opposed to “semantic memory”, which is understood as a more general, abstract kind of memory (such as the memory that one is a mammal)).

counter to some traditions, there is justification for positing such a class completely independently of any armchair arguments. Indeed, current empirical work indicates that philosophers *ought* to consider typing imaginative experiences in this way. I will have more to say about this empirical work in later sections; for now, I offer it as (revocable) warrant for proceeding with the conception of sensory imagination I have suggested.

2. A new, yet familiar, distinction

The main claim to be defended is that we can easily be wrong about what we are imagining, both when remembering our perceptual pasts and when visualizing things unseen. In order to make that case as clearly as possible, it will help to introduce a few familiar and, I think, related ideas: privileged access, success verbs, and intensional (or “referentially opaque”) contexts.

2.1 Privileged Access

I will argue that we can easily be wrong about what we are imagining because we lack *privileged access* to the facts concerning what we are imagining. To say that we have privileged access to some set of facts is to say that a properly functioning introspective faculty is sufficient to give us knowledge of such facts.¹⁴ Note that this is not equivalent to saying that we are infallible concerning the facts to which we have privileged access, since introspection can fail in some cases.

In saying that we *lack* privileged access to what we are imagining, I am contrasting imagination to states like belief and desire, to which we plausibly *do* have privileged access,

¹⁴ Cf. McLaughlin & Tye’s (1998, p. 350) account of privileged access: “*Privileged Access*: When our faculty of introspection is functioning properly, we can know what we are thinking by introspection.”

and assimilating it to a class of activities including seeing and hearing. We plausibly lack privileged access to what we are seeing and hearing because, roughly, knowing such facts requires knowing more than what one seems to be seeing, or what one seems to be hearing (to which we may have privileged access). It requires (roughly) knowing whether the thing one seems to see or hear is really there, which cannot be determined through introspection alone.

Note, however, that just as there is an aspect of visual experience to which we *do* plausibly have privileged access (namely, the intentional content of our experience), so too I will allow that there is a (purely intentional) aspect of imaginative experience to which we have privileged access. The point is simply that we can draw a distinction between what one seems to be imagining, which is fixed by the intentional content of one's experience (and to which we have privileged access), and what one is in fact imagining. Most, if not all, other theories of imagination make no such distinction; usually, what is imagined is simply equated with the representational features of the imaginative experience itself (both for propositional *and* sensory imagination). I argue below (in Section 3) that such a view is highly problematic on its own terms, and that it fails to explain the various intuitive features of imagining outlined in this section.

Finally, some delicacy is required in explicating the privileged access/non-privileged access distinction, since there are uses of 'believe' and 'desire' that create contexts where one arguably lacks privileged access to what one believes or desires. For instance, common usage has it that when we believe something *of* someone or something, or desire something *of* someone or something, we lack privileged access to who or what we believe or desire

something of.¹⁵ Similarly, there are non-propositional uses of ‘believe’—e.g., John believes Sally—that prevent privileged access to what (or who) is believed (if John believes Sally, and Sally is a spy, then John, perhaps unwittingly, believes a spy).

However, these examples pose no problem for the general distinction between facts to which we have privileged access and those to which we do not, so long as we are careful to distinguish propositional belief and desire from belief *of* and desire *of*, and from other non-propositional cases of belief and desire. The argument of this chapter is that ‘imagines’ should be sorted, along with ‘sees’, ‘hears’, and ‘believes of’, into the class of verbs that indicate a context where we lack privileged access to certain relevant facts, not that there is *no* use of ‘believes’ that functions like non-propositional cases of ‘imagine’.

2.2 Success verbs

In comparing imagining to seeing and hearing, I am arguing that ‘imagine’ should be treated as a success-verb. By calling ‘imagine’ a success-verb, I am suggesting that if ‘John imagines *X*’ is true, it follows that certain extra-mental facts relating to this act hold true as well.¹⁶ ‘See’ and ‘hear’ are also success verbs in their non-propositional uses. If ‘Jim sees a banana’ is true, it follows that there is a banana there being seen; and if he hears a canary, it follows that there is a canary there being heard. ‘Remember’ is also a success verb. If Bill remembers that Jill’s birthday is today, then it follows that Jill’s birthday is indeed today (and if Jill’s birthday is not today, then Bill *misremembers* it as today).

¹⁵ For example, if I am unaware that your dog is a Pekinese and believe your dog to be an excellent companion, then I believe *of* a Pekinese that he is an excellent companion, while lacking privileged access to the fact that I believe *of* a Pekinese that he is an excellent companion (and, indeed, without believing *that* a Pekinese is an excellent companion).

¹⁶ The extra-mental facts in question are not of the kind that ground externalist theories of meaning and mental content. The relevant facts must hold whether one interprets the statements that express those facts in an internalist or externalist manner.

It is worth noting that the class of success verbs is not to be identified with the class of *factive* verbs. A verb *V* is factive if *p* follows from the fact that someone *Vs* that *p*. A prime example is ‘knows’: if Jill knows that ice is cold, it follows that ice is cold. Belief is not factive; *p* does not follow from the fact that someone believes that *p*. Notice that the factive/non-factive distinction applies to verbs in propositional contexts (i.e., followed by that-clauses). ‘See’, in its non-propositional uses, cannot sensibly be called ‘factive’ (doing so would require saying that if Jill sees Bob, then Bob). Nevertheless, something *like* factivity attends success verbs; in the case of seeing, if Jill sees Bob, it follows that Bob exists, and (on most accounts) that Bob is impinging on her senses. My suggestion will be that similar things follow if Jill imagines Bob, and that this warrants calling ‘imagines’ a success verb.¹⁷ I will not, however, be arguing that ‘imagines’—in either its propositional or non-propositional uses—is factive.

Where ‘*S*’ is a success verb, it will typically be true that one lacks privileged access to the facts concerning what one is currently *S*-ing.¹⁸ This is because, roughly, it takes more than one’s own mental state to make it true that one is *S*-ing this or that. The world outside one’s skin must also cooperate in some way or other. Indeed, it is no doubt in virtue of these extra-mental conditions *being* extra-mental that introspection is insufficient to provide knowledge of what one is *S*-ing.

¹⁷ I will make some general remarks about what the specific success conditions are in the case of imagination shortly (in Section 2.4), but will leave a full-fledged account for Chapter Two.

¹⁸ N.B.: Lacking privileged access to some set of facts is not the same as being unreliable concerning those facts. Rather, it means only that *introspection* does not provide reliable access to those facts. We lack privileged access to what we are remembering since, in most cases, we need to have knowledge of things other than our own mental states to know whether we are remembering something.

2.3 Intensional Contexts

Another important feature of success verbs—in particular, those relating to perception—is that they do not typically create intensional-with-an-s contexts (or “opaque” contexts). That is, co-extensional terms *can* be substituted *salve veritate* for terms denoting the objects of these acts. If Bill saw your gardener, and if your gardener is also your plumber, then Bill saw your plumber, even if Bill is unaware that your gardener is your plumber. If ‘imagine’ is a success term, we should expect it to not create intensional contexts as well. Intuitively, this is indeed what we find: if Bill is imagining [visualizing] your gardener, then, whether or not Bill believes your gardener is your plumber, Bill is imagining your plumber as well.

Here’s another example. Suppose that Fuji apples are Sally’s favorite fruit. Wishing not to seem extravagant in her tastes, she hides this preference from her husband Ron. She tells Ron that Granny Smith apples are her favorite, and Ron believes her. Nevertheless, if Ron is imagining a Fuji Apple, he is indeed imagining a sample of his wife’s favorite fruit.

‘Imagine’ and ‘see’ differ in this respect from ‘believe’ and ‘suppose’, which are normally held to create intensional contexts. If Ron believes that there is a Fuji apple in his wife’s bag, he does not necessarily also believe that a sample of his wife’s favorite kind of fruit is in her bag. Similarly, if Bill supposes that there is a Fuji apple in his wife’s bag (we ask him to do so), intuitively, it does not follow that he is supposing that her favorite fruit is in her bag. Yet if Ron *sees* a Fuji apple in his wife’s shoulder bag, then he also sees a sample of his wife’s favorite kind of fruit. And if Ron imagines a Fuji apple sitting in his wife’s bag, then *ipso facto* he imagines a sample of her favorite fruit sitting in her bag.

So, there is an intuition that ‘imagine’—at least in its non-propositional uses¹⁹—does not create intensional contexts. This intuition is no doubt intertwined with the intuitions that one can easily be wrong about what one is imagining and that ‘imagine’ is a success-verb; it is hard to know which, if any, is more basic. I have pulled them apart here simply to show that, in the case of imagination, they come together in a way that provides intuitive warrant for grouping imagination with (non-propositional) seeing, hearing, and remembering, instead of with (propositional) believing and desiring.

2.4 *The imaginative experience vs. what is imagined*

Given this grouping, we can expect an analysis of imagination to require distinctions also required for the analysis of these other mental activities. Specifically, in each of the cases where one sees, hears, remembers, or imagines something, we can distinguish between the extra-mental conditions that must hold when one successfully sees, hears, remembers, or imagines that thing, and the purely mental side of the experience, which, on most theories of perception and memory²⁰, can be the same both in successful and unsuccessful instances. For instance, the mental component of seeing a tomato can be exactly the same as the mental component of hallucinating a tomato—they both consist in the representation of a tomato—while only the former “successful” case involves the presence of an actual tomato. Understanding the success conditions for any mental state naturally goes hand in hand with recognizing its *representational (or intentional) content*, for an experience’s representational

¹⁹ I treat propositional imagination separately in Chapter 4 of this dissertation. Note, however, that insofar as some acts of propositional imagination involve the presence of imagery, the present account of non-propositional imagination will be relevant to fully understanding propositional imagination.

²⁰ Notable exceptions are the disjunctive (Martin, 2002) and relational (Campbell, 2002) views of perception.

content is simply what it represents to be the case, and a mental state will be “successful” whenever what it represents to be the case is indeed the case.

It is common to refer to the mental side of a veridical or falsidical act of perception as the *perceptual experience*, and to hold that its nature is (at least partly²¹) determined by its representational content. Two perceptual experiences with the same representational content—that represent the world as being the same way—will involve the same kinds of mental states; a separate question is whether either of the experiences is veridical. I will make a similar suggestion for the case of imagination, appealing to the notion of an *imaginative experience*. Imaginative experiences, like perceptual experiences, are determined by their representational contents. This allows a distinction between the content of an imaginative experience on the one hand, and the issue of whether the success conditions implicit in that content are satisfied, on the other. This in turn allows the word ‘imagine’ to function as a success verb in a way analogous to the way in which ‘see’ is often used. A person will be imagining what he thinks he is imagining when the success conditions for his imaginative experience are met, and will not be imagining what he thinks he is imagining when they are not met. To my knowledge, this is not a distinction that has been made before in the literature on imagination.²² It is commonly assumed that there is no possibility for a gap between what is imaginatively represented and what is in fact imagined, as there is between what is visually represented and what is in fact seen.²³

²¹ Here I am setting aside the question of whether there is *more* to the “mental side” of an experience than its representational content (e.g., its phenomenal character).

²² Though there have of course been other accounts of what being wrong about what one is imagining might consist in (see e.g., Peacocke (1985), discussed below in Section 5.4, and Byrne (2007)).

²³ For instance, in an otherwise insightful discussion of sensory imagination, Noordhof argues that the supposition behind an image together with the intentional properties an image independently possesses account for the total intentional content of an imaginative experience, as a means of establishing that imagining objects does not require imagining perceptual experiences of objects (he argues many people don’t suppose themselves

What precisely *are* the success conditions in the case of imagining? This is the central topic of Chapter Two; addressing it satisfactorily now would take us too far off course. However, in the meantime I will offer the following oversimplified version, to give an idea of where the overall account is headed. I will argue that there are two interrelated forms of imagination, one dedicated to representing what certain things looked like (Memory Imagination), the other to representing what certain things *would* look like (Modal Imagination). This gives rise to two different kinds of correctness conditions. In the memory cases, the imaginative experience is veridical if the thing one is trying to imagine looked roughly the way one's image represents it as having looked (notice that this requires that the thing one is trying to imagine was actually seen at some prior point). For the modal cases, the experience is veridical if the thing one is trying to imagine *would look* roughly the way one's image represents it as potentially looking (which at least requires that the thing one is trying to imagine could possibly exist).

The second key facet of imaginative experiences (both memory and modal) I posit is that they involve two core representational elements: there is the represented *object* at which they aim (i.e., the “thing one is trying to imagine”), and there are the represented *perceptible properties* attributed to that object. Both sets of representational properties are, I argue, fixed by the intentional content of the imaginative experience; and it is up to the will of the imaginer to determine the intentional content of her experiences. What is *not* up to the imaginer is whether the thing she is trying to imagine really looked (or really would look),

to be imagining experiences, and, independently, that images themselves need not represent experiences) (2002, p. 433). But there is a gap in the argument once we accept the distinction I have suggested, for one could allow that the intentional content of imaginative experiences does not contain reference to an experience while leaving open the possibility that *what is imagined* in virtue of having such an experience is in fact a perceptual experience. I point this out not because I think that experiences are indeed the objects of imagination (on this I agree with Noordhof), but rather to show how most discussions simply take for granted that the intentional content of an imaginative experience and what is imagined are somehow the same.

the way she represents it as looking. That is, it is up to the imaginer to determine what she is *trying* to imagine, and what properties she is trying to attribute—but she is only imagining what she is trying to imagine if the object she is “aiming” at really had (or would have) the perceptible properties she represents it as having. This constitutes the advertised distinction between the intentional content of an imaginative experience, and the success conditions in virtue of which one can be said to imagine (or not) the thing one is trying to imagine.

Obviously these claims raise a host of questions—I try to address as many of them as I can in Chapter Two. For now I hope this suffices to show what theory that posits success conditions for imagination might look like in its general outlines.

3. What the representational content of imagination is not

As a way of supporting my claims concerning what success conditions for imagination are, let me say a little about what they are *not*. Contrary to a popular view, imaginative experiences do not (however “weakly”) represent the *presence* of some thing or other. This may seem intuitively obvious—no person in his right mind comes to believe that a tiger is present based on his having imagined a tiger. However (sensory) imagination is often defined as “quasi-perceptual experience in the absence of appropriate stimuli” (Gendler, 2005; Thomas, 2007).²⁴ To say that a state has occurred in the absence of “appropriate” stimuli, however, is to suggest that the state is non-veridical in some sense—

²⁴ The definition of sensory imagination as “quasi-perceptual experience in the absence of appropriate stimuli,” is vexed even overlooking the problem I discuss above. For if imagination is *quasi*-perceptual to begin with, it is redundant (or simply false) to add that it occurs in the absence of appropriate stimuli. And, presumably, imagination is not simply *perceptual* experience that occurs in the absence of appropriate stimuli, since we normally reserve the title of “illusory” or “hallucinatory” experience for that sort of thing (and, not to mention, illusory and hallucinatory experiences play very different cognitive roles than do imaginative experiences). Thus, such “definitions” of sensory imagination hang everything on the word ‘quasi’ and so tell us nothing that will differentiate it from perception or hallucination until the import of ‘quasi’ is explained.

that it misrepresents. But if, when I imagine a flying cow, my imaginative state does not purport to represent an actual nearby flying cow, why think that it has occurred in the absence of appropriate stimuli?²⁵ On the face of it, my intention or desire to imagine a flying cow is a perfectly *appropriate* cause of a flying cow image, so an appeal to the inappropriateness of its cause cannot be used to differentiate images from percepts. The case is even clearer for acts of memory imagination: when I imagine my childhood home, I do not, in the absence of appropriate stimuli, represent it as present before me. I do not even represent it as still existing; if it has been torn down, my experience need not have misrepresented it in any way.

Nevertheless, the view that imagination represents objects as existing before us—perhaps just not as “vividly” as perception—remains widespread.²⁶ In arguing that sensory imagining does not involve the appearance of possibilities (contrary to my position in Chapters Two and Three), Alex Byrne claims:

If sensuous imagination involves the appearance of anything, it’s the appearance of *actuality*, not possibility. What “appears to be so”, when one imagines a purple polar bear, is that purple polar bears exist, not (merely) that they could have existed. (Of course, one will not believe that things are as they appear). (2007, p. 134, my emphasis).

According to Byrne, our imaginative experiences represent various things as actually existing (though he doesn’t say *where*, exactly, these things are represented as being). Fortunately, we are not so easily duped—we “will not believe that things are as they appear.”

²⁵ Cf. Searle (1983, p. 18): “It is not a failure of a state of imagination that nothing in the world corresponds to it.”

²⁶ See, e.g., Goldman (2006a, pp. 48–49): “The real experience and the [sensory]-imagining of it (when accurate) are intrinsically pretty similar...being the product of the will versus the environment is only an extrinsic difference...the different epistemological status of images and percepts is a direct result of their extrinsic [and not intrinsic] differences.” On the plausible assumption that representational properties are intrinsic properties of an experience, Goldman is committed to sensory imaginings being (in general) *non-veridical*.

Byrne thinks that reinforcement of this claim is available through consideration of the “Perky effect”, which he glosses as “mistaking perception for sensuous imagination.” The Perky effect is a piece of philosophical folklore that often arises in comparisons of imagination to perception, so it deserves some special attention here.

The Perky Effect.

Perky (1910) reports an experiment that has no doubt attracted more philosophical attention than any other empirical study involving silent movies. Through a painstaking process described in detail (1910, p. 428-431), Perky and her colleagues were able to create conditions where subjects regularly mistook dimly and hazily projected images on a back-lit screen for mental images of their own creation. A crucial aspect of the experiment design involved concealing the fact that subjects were looking at a (glass) surface that could have images projected onto it from the other side.²⁷ A number of philosophers, Byrne (2007) included, have taken Perky’s results as an indication that images have the same representational properties as percepts—namely, that they represent objects with various properties *as present before one*.

This seems to me a rash conclusion. The very fact that such special circumstances need to be put in place before any such confusion occurs points strongly in the opposite direction. As Perky reports, when there was any sort of “technical error” (a camera wobble, an edge of light showing in the corner of the screen, etc.), the subjects immediately

²⁷ The very novelty of projected images in 1910 no doubt played a role in convincing subjects they were imagining the figures, and also likely accounts for the difficulty modern experimenters have had in duplicating Perky’s findings (see, e.g., Segal (1971)). However, Segal attributes her own inability to duplicate Perky’s results to the “suspicious, pragmatic students who populated our campuses [and served as test subjects] in the late 1950’s and early 1960’s, who had not been ‘turned on’” (1971, p. 77).

determined that they were not imagining the figure in view. One could more easily argue that Perky's experiment shows that imagining is very *different* from perceiving (even different from *dimly* perceiving), given the great lengths to which one must go in order to make someone confuse the two. At any rate, the form of argument is no good: the fact that subjects may, under extraordinary circumstances, mistake an X for a Y, is not good evidence at all that Xs and Ys are in essence the same (or even that they *look* the same). What's more, the confusion between imagination and perception can easily be attributed to shared representational properties *other than* those which, in perception, account for a thing being represented *as present*.

Finally, there is some dispute over whether Perky's results have been duplicated, and hence over their ultimate legitimacy. Insofar as there is room for disagreement, it appears terminological in origin. Byrne and other philosophers use 'the Perky effect' to refer to the mistaking of percepts for mental images, while psychologists (Craver-Lemley & Reeves, 1992; Segal, 1971) generally use 'the Perky effect' to refer to the *interference* of visual imagery with concurrent visual perception tasks, such as when active image formation causes subjects not to notice perceptible visual cues they would otherwise notice. The latter phenomenon is well established, and can be explained by the fact that imagery and visual perception draw upon common neural mechanisms. However there have been few if any duplications of the "Perky effect" as philosophers understand it. The closest are described by Segal (1971), who notes that she was forced to give subjects a placebo, which they were told would cause them to "relax", before *some* would claim that projected images were their own

mental images.²⁸ Clearly, none of these findings give us serious reason to think that sensory imagination is, in its representational properties, identical, or even very similar, to perception.

Alvin Goldman (2006a, 2006b) also argues that sensory imaginings involve representing various objects as present, in just the way the perceptual experiences typically represent objects as being present before one (I will call such views “as present” views of imagination). Sensory imagination is a form of what he calls ‘enactment imagination’ (or E-imagination), where one “creates, though imagination, perception-like states” (2006b, p. 42). In what way is sensory imagination “perception-like”? Goldman thinks that sensory imaginings and percepts “are intrinsically pretty similar”, and suggests that the differences typically noted between the two are “a direct result of their extrinsic [and not intrinsic] differences” (p. 48-49). On the plausible assumption that representational properties—in particular, the representational properties in virtue of which something is represented as present before one—are intrinsic properties of an experience, Goldman is committed to the view that sensory imagination, like perception, represents things as present.

If imaginative experiences have the same content as perceptual ones, why do they not lead to the same kinds of behavior? In particular, why don’t they cause us to think that the things represented are present before us? In response to this rather fundamental query, Goldman notes that perceptual experiences themselves do not always lead to behavior appropriate to the thing represented being true. Just as forming a visual image of the surf

²⁸ Segal adds that some patients apparently viewed the placebo as a kind of hallucinogen, as they would continue to claim the projected images were mental images even once the intensity of the stimulus was raised well above threshold (1971, p. 77).

does not lead one to reach for one's surf board and begin paddling, so too an instance of seeing the surf may lead to no such behavior because:

Perhaps there is no surfboard available; perhaps one sees the surf on television; perhaps one's leg is in a cast that makes surfing impossible. Still, similar modes of mental action preparation would flow both from seeing the surf and from having a visual image of the surf (p. 47).

Overlooking the fact that the visual experience of seeing the surf on television is not all that much like actually seeing the surf, Goldman fails to address the main dissimilarity between imagination and perception. The key causal effect of percepts is typically a belief in the presence of the thing represented; they no doubt have this effect in virtue of their own representational properties. Even if your surfboard is cracked and your leg is broken, a visual experience as of the surf will, in the absence of conflicting information, lead you to believe that the surf is there before you. Imaginings of the surf have no such effects in normal circumstances. So why think that they, too, represent various objects as present?

Goldman's main argument seems to rest on the neurological and behavioral evidence linking imagination to perception. He discusses a range of empirical studies that show a significant overlap in the neural regions and cognitive mechanisms at work in both perception and imagination (I discuss similar data in Chapter Five), concluding that "E-imaginative events have substantial similarities—including functional similarities—to their actual counterparts" (p. 44). But one can easily accept that there are substantial similarities between imagining and perceiving (as I do) without concluding that one of those similarities is that they both represent things as present. After all, the overlap is far from *complete* between the processes used in vision and those used in imagination (Kosslyn, Thompson, &

Alpert, 1997). One should *always* be cautious in inferring shared *psychological-level* properties (i.e., representational or intentional properties) directly from shared neurological properties—even more so when the overlap is only partial.

Goldman's other main argument is that imaginative states often give rise to affective responses similar to those of perception (pg. 49-54). Again, everything rides on how strongly we interpret "similar". If imaginings reliably gave rise to the *very same* affective responses that their counterpart perceptions do, this would be good reason to suppose they have the same representational properties. But he offers no evidence of *that* kind. Rather, the work he discusses (e.g., Harris, 2000) reveals some similarities in affective response between what we imagine and what we perceive, some of the time.²⁹ This warrants concluding at most that imagination and perception share *some* representational properties. Below I offer empirical evidence supporting the contrary view that sensory imagination represents how things looked, or would look—not how they *currently are* before one.

So far my arguments against the *as present* view have not relied on any claims about whether we can, or cannot, be wrong about what we are imagining. But assuming now that we can indeed be wrong in the ways I have suggested, we can see that the *as present* view of imagination is poorly suited to make sense of this phenomenon. Like the view of sensory imagination as "quasi-perceptual experience in the absence of appropriate stimuli," the *as present* view of imagination implicitly assumes that all imagination is non-veridical—that it represents the world to be a way that it is not. On a simulationist view like Goldman's, we intentionally form these non-veridical representations as a means of forming a judgment

²⁹ See my more extended discussion of affective response to imagined scenarios in Chapter Four.

about what someone else is thinking, or in order to plan some action of our own, or for some other task. Goldman might emphasize this in response to the above criticism; we do not act as though an x is present when we (imaginatively) represent one as present because we know that we have willfully created this non-veridical representation. If this is one's view of imagination, then to imagine is (roughly) to *willfully perceptually misrepresent* some situation. This allows for a story about how it is that imagination can misrepresent, for it *always* misrepresents by its very nature (absent extraordinary circumstances).

But this means that, whether one thinks that propositions or worldly objects are the things we imagine, we will (again, absent very odd circumstances) always know what we are imagining. For if what we imagine is the propositional content of a certain mental state, introspection should be able to reliably tell us what we are imagining. And, alternatively, if what we are imagining is the worldly object that the experience purports to represent, then, assuming privileged access to our intentions, we will always know that there is no such object present, since imagining amounts to willfully misrepresenting that object (presumably one would successfully imagine an object by willfully perceptually misrepresenting it). So it is unclear how someone could be wrong about what he is imagining on such a view, regardless of how one understands the objects of imagination.

Of course, instead of trying to accommodate the intuition that we can be wrong about what we are imagining, a defender of the *as present* view can simply try to explain it away, showing how it is somehow mistaken. I look at the prospects for doing so in Section 4, and argue it cannot be done.

Fitless Imagination

In lieu of claiming that imagination represents things as present, some have claimed that imagination is “fitless” (i.e., lacking in a direction of fit), and therefore without genuine correctness conditions (Lormand, 2007, fn. 15; Searle, 1983). Searle proposes that such states are merely “as if” they have correctness conditions. More recently, Colin McGinn has put the point this way:

Percepts supply (defeasible) reason to believe; they insist on their own veracity. But images do not invite belief in this way; they do not purport to tell us how the world is. They are neutral about reality...In the typical case we have no tendency to believe what our images represent: if I form an image of a tiger in my room, I am under no impulsion to believe there is a tiger in my room...Not only does the image not tell us about the external world in the way the percept does; it does not even try to: it is not in this line of business (2004, p. 21).

McGinn sensibly wishes to avoid the claim that images purport to tell us about what is the case in our immediate environment, partly as a means of distinguishing images from percepts. Note, however, the following instability in his account. He puts his point that “images do not purport to tell us how the world is” by saying that we have “no tendency to believe what our images represent.” This involves a *prima facie* contradiction, since claiming we have no tendency to believe what our image represent is to say that images *do* represent the presence of this or that leaping tiger after all, and that we simply know better than to trust such non-veridical states. This is to fall into the very trap McGinn was seeking to avoid—it is to hold that images *purport* to tell us that certain things are in fact present in the same way that percepts do.

Most likely what McGinn and Searle mean to suggest is instead that forming an image is akin to *merely entertaining* a content, without taking *any* attitude, assertoric or desirative, toward the content. On this view, my mental image of a tiger does not represent a

tiger as present; it merely represents a tiger (some tiger, somewhere?); thus, there is no need to *disbelieve* what it represents.

There are two main problems with this way of understanding imagistic experience—one concerns the coherence of the general notion of “merely entertaining” a content, the other is specific to the case of imagination. First, the notion of *merely* entertaining a content (or image), without taking any attitude at all towards it, is more obscure than I think is normally recognized. There is a sense of ‘entertaining’ according to which all kinds of thoughts—beliefs, desires, regrets, etc.—count as particular *ways* of entertaining a proposition. But this is to use ‘entertaining’ as a general category term like ‘animal’; there are many *kinds* of animals, but there are no *mere* animals—no animals that are of no particular kind or other. So what would it be to *merely* entertain a thought or image, while taking no attitude toward it? Is it to sort of wink at the content, avoiding eye contact? Is it to simply hold the proposition “present before one’s mind”? What does that consist in? How do we know when we are doing it? Is it supposed to be introspectively clear that I can merely entertain a proposition, while having no attitude at all toward it?

When I try to *merely* entertain a proposition—the proposition that, for instance, goats are made of cheese—by, say, “holding it present before my mind”, I inevitably find myself having various beliefs and posing various questions to myself, all related in some way or other to the claim that goats are made of cheese. There is, for instance, puzzlement about what one who said “goats are cheese” might metaphorically mean; there are thoughts about what a piece of cheese in the shape of a goat would look like; there are thoughts about various other qualities I believe goats and cheese to have. Yet all of these thoughts are perfectly normal assertoric or erotetic thought episodes—why, in addition, should we

countenance a floating, *merely* entertained, proposition that is also mentally present? What if I “say”, in inner speech, “goats are made of cheese.” Have I thereby (merely) entertained the proposition that goats are made of cheese? But isn’t subvocalizing something I could do even if I had no idea what ‘goats are made of cheese’ is supposed to mean (if I were, for instance, a non-English-speaking tourist, who overheard a philosopher say ‘goats are made of cheese’ on the subway, and then repeated the phrase subvocally)? Keep in mind, it is one thing to think *about* the proposition that goats are made of cheese, forming beliefs and questions concerning *it*, which is what I take myself to be doing when told to “merely entertain” the proposition; it is quite another to think *with* that proposition, while taking no attitude at all toward it. The latter activity I do not understand—not *really*.

I will have more to say about this issue in Chapter Four on propositional imagination. That is where I will consider accounts that analyze propositional imagination (and supposition) as involving a distinctive cognitive attitude and corresponding aspects of cognitive architecture, which might be the very thing one might wish to appeal to in elucidating the notion of the “mere” entertaining of a thought. For now, suffice it to say that I find the notion of merely entertaining a thought to be far more obscure than most philosophers normally let on. Quite generally, it is the intentional mode taken towards a proposition that enables us to understand why a certain mental state plays the role it does—propositions cannot predictably influence behavior all by themselves.³⁰ A desire that the Cubs win the World Series, and a belief that the Cubs won the world series have clear—and very different—psychological implications; the mere proposition *that the Cubs won the*

³⁰ The best case for the mere entertaining of an image influencing behavior comes from affective responses to, for example, the request to consider a vomit sandwich. These responses do not aid in the systematic recall of information, and in action planning, in the ways that so much of imagination does.

World Series by itself has no clear cognitive role. To play any sort of predictable cognitive role, propositions must be paired with an intentional attitude, one which gives the mental state satisfaction conditions of various kinds. Merely claiming, as Searle does, that imagination is “as if” it has correctness conditions is to dodge the crucial question and, at any rate, provides no insight into the cognitive role of imaginative experiences.

This sort of general worry aside, another serious problem with holding that imagining is the “mere” entertaining of thoughts and images is that lots of empirical research (not to mention intuition) shows that imagining influences behavior and reasoning in profound ways, and it is mysterious how it could reliably do so without (veridically) representing conditions in the world. I will take a moment now to discuss some of this work in detail, as its appreciation is important to understanding the success conditions of imaginative experience.

3.1 Some empirical considerations

There is plenty of anecdotal and intuitive evidence that imagery plays a key role in representing the world—that it has satisfaction conditions in more than an “as if” manner. When people are asked about structural, spatial, or visual details of objects they are familiar with—the number of doorways in one’s home, or the color of one’s first car, for instance—they typically report forming images to do so. These imaginative acts plausibly represent real objects as having certain properties, and surely misrepresent in cases where the objects lack (or *lacked*) those properties. Intuitively, similar uses of imagery occur in cases where one imagines what certain things would look like if they were seen (and not simply how certain things looked in the past). Athletes and performers form images to imagine what it will be like to complete certain desired actions, and have been shown to improve

performance on those tasks simply through imagining them (Janssen & Sheikh, 1994; Suinn, 1994).³¹ Yet similar cases of imagining what something *would* look like *if* seen (what I will call ‘modal imagination’) occurs in completely quotidian contexts. If I see some chairs in a store I am considering buying for my kitchen table, I might imagine the table with chairs of that sort around it in order to determine how many would fit. This sort of imaginative act does not involve remembering my table with those chairs around it, but nevertheless still seems to have correctness conditions; if I imagine six chairs around the table, but only four would fit, then the imaginative act misrepresents how many chairs would fit.

Yet these are still mainly intuition-based cases; to put the claim that imaginative experience has genuine correctness conditions on firmer ground, it will help to look at some cases from the empirical literature. Rubin and Greenberg (1998, 2003) and Greenberg *et al.* (2005) describe a fascinating form of amnesia—called visual memory-deficit amnesia—that results from damage to visual centers of the brain (primarily the occipital lobe) that underlie the ability to generate visual imagery (Kosslyn, et al., 1997; Slotnick, Thompson, & Kosslyn, 2005). While others have shown that an ability to generate visual imagery is not *sufficient* for autobiographical episodic memory (AEM)³² (damage to centers in the medial temporal lobe involved in memory retrieval and organization often causes amnesia as well) Rubin and Greenberg tested whether visual imagery might nevertheless be *necessary* for AEM. They

³¹ Suinn (1994) quotes golf legend Jack Nicklaus:

I never hit a shot, not even in practice, without having a very sharp, in-focus picture of it in my head....First I ‘see’ the ball where I want it to finish, nice and white and sitting up high on bright green grass. Then the scene quickly changes and I ‘see’ the ball going there: its path, trajectory and shape, even its behavior on landing. Then there is a sort of a fade-out, and the next scene shows me making the kind of swing that will turn the images into reality (1994, p. 23).

³² Autobiographical episodic memories are memories of one’s own past that are specific with regard to time and place—a memory of a family reunion dinner in South Carolina, for example. Psychologists typically distinguish these kinds of memories from “semantic” memories, which are more general in nature (for instance, that one is an American (autobiographical semantic), or that whales are mammals (non-autobiographical semantic)). Most forms of amnesia consist primarily of a loss of autobiographical episodic memories.

reviewed the literature for cases of visual memory and imagery loss, to see whether deficits in visual memory and visual imagery might correlate with a broader loss in autobiographical memory.

They used the following criteria (from a study by Farah (1984)) to define visual memory loss:

- 1) the patient is able through vision to detect, draw, or describe an object present before them (establishes the patient is not merely blind),
- 2) the patient is not able to name or give the function of the object (criteria 1 and 2 together define the patient as a visual agnostic), and
- 3) the patient is not able to draw an object from memory, describe its visual characteristics from memory, or detect its visual image upon introspection (1998, p. 5414).³³

While subjects were selected according to these merely behavioral criteria, it is worth noting that they all had cortical damage consistent with a visual deficit.

Despite the inability of these patients to name and state the function of objects they could see and draw, they could readily recognize the same objects once the objects were placed in their hands. Nevertheless, the seemingly quarantined visual-perceptual-deficits were correlated in all eleven subjects with a much broader amnesia concerning general personal histories, both in retrograde (before trauma) and anterograde (after trauma) directions. Large amounts of information about their personal pasts (and, in some cases, nearly *all* such information) was simply no longer available once their ability to form images

³³ As Rubin and Greenberg note, it is not enough simply to study subjects who report an inability to form images, since brain-damaged patients' introspective reports may be especially unreliable, and some individuals with no known neurological damage report an inability to form visual images.

was impaired. While the cortical damage in some of these patients extended beyond solely visual areas of the brain, at least two of them had brain lesions limited exclusively to visual cortex. The take home point for the purposes of this chapter is that the ability to form visual images underlies a vast range of representational abilities, extending well beyond the mere ability to think about the visual characteristics of things. One's memory of one's own personal past—and *ipso facto* one's sense of self—is deeply connected to the ability to generate imagery.³⁴

While the Rubin and Greenberg study showed an imagery deficit to be linked with profound amnesia—that is, with deficits in representing events from one's personal past—the already mentioned work by Schacter *et al* (2007) and Buckner and Carroll (2007) establishes a further neurological link between the mechanisms that underwrite autobiographical episodic memory and those that enable the imagination of novel or fantastical scenarios. Buckner and Carroll discuss a variety of studies of amnesiacs who are unable or unwilling to contemplate taking part in future activities, as well as imaging studies that reveal activation

³⁴ It may seem puzzling that memory could rely so heavily on visual abilities, given that the blind (who presumably do not form visual images) do not show any special memory deficits. Rubin and Greenberg's claim that AEM requires visual imagery is actually qualified as follows: *if* memories were originally encoded in a visual modality, then visual imagery will be necessary for their recall. Rubin and Greenberg posit that the blind find other ways to encode memories (perhaps co-opting some of the same neural networks), and draw on those abilities in AEM. Indeed, the ability of the blind to develop distinctive modes of encoding memory has relevance to two important predictions confirmed by Rubin and Greenberg's studies. First, they predicted that the amnesia of patients who suffered damage to primarily visual centers would be more severe in the retrograde (before trauma) direction than anterograde (after trauma) direction, since new strategies for encoding and retrieval could presumably develop post-trauma, just as the blind develop alternative ways to store memories. However, visually-stored memories would, on their hypothesis, not be retrievable once the relevant visual centers no longer function. This is an interesting prediction since it is uncommon for severe retrograde amnesia to be coupled with only mild anterograde amnesia (Kapur, 1993). The unusual predicted pattern was indeed found in five of the seven patients where relevant data was available.

A second prediction was that there would be little to no temporal gradation in the severity of the retrograde amnesia. This prediction also runs counter to what is normally seen in amnesia, as the memory deficit is usually less severe for memories formed further in the past (childhood memories, for instance, are normally fairly well intact). Again, the data, while not univocal, supported their hypothesis: four of the six patients for whom relevant data was available showed no temporal gradient in their amnesia; even childhood memories were not spared for these patients.

in the same core brain network during AEM and the imagination of future experiences (p. 51-53). Schacter *et al* discuss similar studies, arguing that available data “converges on the conclusion that the process of imagining or simulating future events depends on many of the same neural processes that are involved in episodic remembering” (p. 659).

When we combine this neurological and behavioral links between AEM and future—“prospective”—planning with the key role that Rubin and Greenberg’s work suggests for imagery and visual memory in AEM, there is sound empirical reason to think that much planning and strategizing about one’s personal future relies upon imagery as well. If that is right—as is certainly supported by the intuitive considerations with which I began this section—then, far from being “fitless”, as some have supposed, imagination is intimately tied to a variety of representational tasks. Most saliently, these include representing how things were in one’s perceptual past, and how things will (or might) be in one’s perceptual future. Moreover, these are representational tasks that clearly allow for the possibility of *misrepresentation*. In the retrodictive and prospective cases respectively, one’s imaginative states can misrepresent either how things were or would be. When they do, we will be wrong about what we are imagining.

A simulationist of Goldmans’ stripe certainly would not deny that imagery plays a key role in various representational tasks. The main difference is that, in focusing so intently on the similarity of imagination to perception, the simulationist view ends up analyzing imagination as a kind of perceptual misrepresenting that can serve as input to some other inference-making mechanisms. I suggest that, in order to remember our past experiences, or even to for a judgment about someone else’s current experience, we do not need to misrepresent the current situation and then *transfer* that to a judgment about the past, or

about another person. It is rather in the nature of imagination to directly represent past events with the use of imagery, or to directly represent what someone else's current, or your own future experience might be like.

I am thus suggesting a difference in how we cut things up at the psychological level—a suggestion warranted by the dissociations we seen in Rubin and Greenberg's analysis of visual deficit amnesia. The patients they studied were able to draw and describe objects through sight (though they could not usually say what *type* of object they were seeing). Their visual-perceptual faculties were largely intact. And yet their abilities to recall, through imagination, their own pasts were severely—often completely—impaired. While Rubin and Greenberg did not assess whether similar deficits existed in the prospective direction (imagining future experiences) Schacter *et al* (2007), Addis *et al.* (2007) and Hassabis *et al.* (2007) found that such deficits are indeed mirrored in their (in)abilities to imagine the future as well. This suggests that, instead of viewing imagination as an “offline” running of perceptual mechanisms, we understand the imagination as a distinct faculty whose role it is to enable “episodic thought” in two temporal directions (as Schacter explicitly suggests). This is a faculty that draws heavily on long term memory and on visual-perceptual mechanisms, some of which overlap with those at work in perception. Yet, as we see with the patients Rubin and Greenberg discuss, an organism can still successfully perceptually navigate its environment with a complete loss of its imaginative capacities. This shows that, as has sometimes been supposed, non-human animals with acute visual perception may nevertheless lack the ability to visualize the past and future. This again weighs in favor of

seeing imagination as a distinctive ability, with distinctive representational contents, and not as a mere offshoot of perception.³⁵

In closing this section, I will briefly consider two potential worries. One may object that all this only shows that imagination, be it retrospective or prospective, *often does* have an assertoric element in virtue of which it represents ways the world was, is, or would be—not that it *always must* have this element. True enough, but I seek to analyze the central cases of imaginative experience—the ones that contribute most to our cognitive lives. If these cases are representational in the way I have suggested, and if there is no compelling reason to think that other outlying cases *must* be different, then we have warrant to tentatively conclude that all imagination is representational in one of the ways I have suggested.

That imagination should have contents of the kinds suggested—that it should involve so much *commitment* on the part of the imaginer—may seem counterintuitive (this is the second worry). But note that, while there may be acts of imagination where we do not feel “committed” to the veracity of our imaginative act (maybe we don’t *really care* if a speckled hen would look like thus and such) this does not show that the imaginative experience itself lacks correctness conditions. The correctness conditions are posited to account for the central cases of memory and planning where we *do care* about the veracity of our imaginative experiences, and are assumed to be present even in the odd cases where we do not care since *it is in the nature of such states to have such correctness conditions*. So, just as we can perceptually represent that one line is longer than another while not believing it to be

³⁵ The simulationist can of course respond that what non-human animals (and Rubin and Greenberg’s patients) lack is the ability to take their perceptual systems “offline”.

(as in the Muller-Lyer illusion), because it is *in the nature of* perceptual experience to represent how the world is, similarly one's imaginative experience can have correctness conditions even when one has little interest in whether those conditions are met. One can “disregard” the matter of their satisfaction in the same way one can “disregard” the matter of one's perceptual experience being veridical or not without this changing the representational properties that those states in fact have.

4. Imagination and Intent – explaining away the privileged access intuition

So far I have sketched a theory of imagination that takes seriously the success-logic of ‘imagine’. Part of making that case involves an appeal to empirical work that puts imagination in the business of representing past and future conditions. Note that even when counterfactuals are part of the content, the experience represents the counterfactuals as being true of *this world*; it is a fact about this world that if I were to see a cube (in optimal conditions) it would look a certain way. I will fill in the details of this approach to imaginative content in Chapter Two. First I want to look more closely at the main intuition that clashes with the account of imagination I've been developing; this is the intuition that we have privileged access to what we are imagining.

I will look at two main ways of motivating—and defusing—the privileged access intuition. One traces our privileged access to the role that intentions play in determining what we imagine. Some argue that, because our intentions determine what we are imagining, and because we have privileged access to our intentions, this guarantees privileged access to what we are imagining. I call this the ‘objection from intent’.

The source of the other is more nebulous; it is generally advanced by those who allow that imagination can occur without one's intent (such as when a song gets stuck in one's head, or when upsetting images flash before one's mind). These theorists argue that, despite imagination often occurring against our wills, whenever we introspect we can clearly determine what we are imagining. I call this the 'phenomenological objection' to my account, for lack of a better term. I begin with the objection from intent.

4.1 The objection from intent

It is undeniable that our intentions typically play a role in imagination that they do not play in perception. This suggests, to many, a similar difference in our epistemic access to the objects of the respective acts. However, I think the latter claim is unwarranted. We can accept that our intentions usually play a role in determining what we imagine, while resisting the conclusion that this gives us privileged access to what we are imagining.

Let's begin with the sort of example I've been considering all along: I mistakenly believe the man I saw yesterday, your gardener, to be your podiatrist. Today you ask me to imagine your podiatrist. I fully intend to do so, but, on the account I've been giving, fail. Instead I imagine the man I saw yesterday, who is a gardener. If that is the whole story, then one's intent clearly does not fix what is imagined, and privileged access to one's intentions does not guarantee that one has imagined what one intended.

Now, the defender of the objection from intent wants to reply that I have misdescribed the situation. My intent all along was to imagine *the man I saw yesterday*, which I succeeded in. If we simply get clear on what my actual intention was, the objection

goes, it is clear that my intention *succeeded*. Moreover, this is an intention to which I had privileged access.

The problem with this response is that, if we hold that our intentions always succeed in fixing what is imagined, we must deny that, in the case described, I really intended to imagine your podiatrist. And if we deny this—even though I sincerely tell you that I intended to imagine your podiatrist—then we are forced to conclude that I do *not* have privileged access to my intentions (since I sincerely claim to have had an intention that I supposedly did not have, and am not suffering a cognitive or memory deficit). And if I do not have privileged access to my intentions, then the objection from intent loses its footing; the putative fact that intentions determine what is imagined will result in no special epistemic access to what is imagined.

The issue that must be confronted by both sides is that we can unknowingly have incompatible intentions. I really did, it seems, intend to imagine the person to whom you referred, namely, *your podiatrist* (I didn't intend to *disobey* you, after all); and yet it is equally plausible that I intended to imagine the man I saw yesterday. I can unknowingly have incompatible intentions in the same way I can unknowingly have incompatible beliefs. I can mistakenly think the two descriptions, "the man I saw yesterday" and "your podiatrist" refer to the same person, and so simultaneously intend to imagine both (just as I can simultaneously believe that I am seeing the Morning Star and believe that I am seeing Mars, if I mistakenly believe that the Morning Star is Mars). Introspection alone cannot tell me which—if any—of the descriptions I have in mind correctly applies to the thing I am imagining.

The result is that, even if my intent does normally play a *role* in determining what is imagined, and even if I have privileged access to my own intentions, this is not enough to give me privileged access to the facts concerning what I imagined. Oedipus does not intend to kill his father or to marry his mother—if he did, we would evaluate his behavior differently. Yet, when he imagines the man he slay lying in the road, he imagines his father, whether he intends to or not.

4.2 Tye's proposal: the phenomenological objection

The fact that we can falsely believe two descriptions apply to the same person or object will not be enough to convince some that we lack privileged access to what we are imagining. For one may reply that certain descriptions under which things are imagined are *special*, and that we can know solely through introspection that *these* descriptions correctly apply to the things we imagine. I will call this lingering worry the “phenomenological objection” to my dismissal of the privileged access intuition. Regardless of the role of intent in imagining, it just seems that, when I introspect my imaginative experience, I am sure to be right about certain properties of what I am imagining, even if for others I am easily susceptible to error. For instance, in the case of the gardener who claimed to be your podiatrist, I took myself to be imagining a man with a peculiar moustache, and, intuitively, was correct so far as that went. Did I not at least have privileged access to the fact that I was imagining *a mustached man*?

By making the object of my imagination suitably vague—that is, specific *in only certain respects*—perhaps I can ensure that I have privileged access to what I am imagining. This seems to be Michael Tye's view (1991, Ch. 2). Tye argues that other writers, most

notably Gilbert Ryle, have been incorrect in their claims that we can always identify the objects of our imaginings. Surely, Tye observes, one can be shown an object of a kind that one cannot identify, and nevertheless imagine that very object, failing to identify it as the kind of thing it is. Nevertheless, he thinks there is something right about Ryle's infallibility claim:

“Where Ryle goes wrong, I suggest, is in ignoring the distinction between having a mental image of one specific object (in which case, there must really be an object in the world which one is imagining) and having a mental image of an object of kind *F* (though no one object of kind *F* in particular). In the latter case, it seems plausible to hold that one must identify what is imagined as being an *F*. In the former case, however, there need be no such identification. What is true here for imaging is true also for seeing, it seems to me. The analogy to the latter case is seeing *that* there is an *F*; the analogy to the former is seeing an *F*.” (1991, p. 21).

Tye thinks that if I set out to imagine a particular object in the world, I can succeed in doing so while failing to identify what it is that is imagined (i.e., failing to identify it as the kind of thing it is). What I cannot do, Tye says, is imagine an unspecified thing of a certain kind (i.e., “some *X* of kind *F*”) without identifying it as being of that kind.

If Tye is correct, then, in order to avoid being wrong about what I am imagining (or, in my terms, in order to have privileged access to what I am imagining), I simply must avoid specifying any particular object as the object of my imagining. Instead I should hold that I am imagining *a* (but no particular) mustached man, or *a* (but no particular) speckled hen. For as soon as a worldly particular is specified as the object of my imagining—for instance, *the* speckled hen I saw yesterday—I can easily be wrong about what I am imagining. The actual object trumps, so to speak, my belief about the content of my imaginative experience. It may be a certain speckled rooster, not a hen, that I am imagining, just as it was your gardener who I was imagining earlier, not your podiatrist.

While there is an intuition worth considering here, I do not think that Tye has correctly expressed it. For it seems I can *easily* imagine an unspecified thing of kind *X* without identifying it as being of that kind. If Fuji Apples are Sally’s favorite kind of fruit, and I imagine an unspecified Fuji Apple, I am also imagining an unspecified sample of Sally’s favorite kind of fruit, whether I can identify it as such or not. Conversely, if I can identify Sally’s favorite kind of fruit by sight, but do not know them to be Fuji Apples, I could still imagine some, but no particular, Fuji Apple by imagining some, but no particular, sample of Sally’s favorite kind of fruit. Similarly, a person who believes whales are fish is imagining a mammal when he imagines an unspecified whale, even though he cannot identify the object of his imagination as such.

The analogy that Tye seeks between imagining an unspecified object of kind *F* and ‘seeing that’ something is an *F* is also mistaken. He proposes that imagining an unspecified thing of kind *X* and seeing *that* something is an *X* are similar—presumably because both require exercise of the concept *X*—while imagining a worldly particular that is of kind *X* and seeing an object of kind *X* are similar because they do not require exercising the concept *X*. But, again, I can easily imagine an object—specified or unspecified—of kind *F* (e.g., where *F* = Fuji Apple) without even possessing the concept *F*, while, intuitively, it remains true that one cannot see *that* something is an *F* without exercising the concept *F*.³⁶

³⁶ I think the source of Tye’s confusion ultimately lies in his taking images as the objects of introspection (he speaks of “inspecting images”), instead of states in virtue of which we imagine objects. Note the instability in his picture: on the one hand, a certain image can be *of* an object that I cannot identify, suggesting that imagining concerns worldly objects. But Tye does not want to say that we can be completely wrong about what our image represents because, when we *inspect an image*, it seems we cannot be wrong at least about the kind of thing represented (so long as the external world does not interfere). But when we “inspect an image” we are not imagining—rather we are involved in a higher-order awareness of a certain first order imagistic mental representation. Imagining does not involve inspecting images, it involves *imagining things*, just as seeing does not involve inspecting visual experiences—it involves *seeing things*.

A better way of voicing the intuition behind Tye's claim, I think, is to admit that we are sometimes wrong concerning *aspects* of what is imagined, while holding that we nevertheless always have available the resources to redescribe the object of our imagining in a way that renders us correct concerning what we are imagining (or, at any rate, that there are always aspects of what we are imagining to which we have privileged access). The likely candidates for these aspects are, in the case of visual imagination, the visually observable properties of the thing imagined.

But, so long as the objects of imagination are *objects* (and their properties), and not images, then it seems we lack privileged access to their perceptible properties as well. Suppose that last week I saw a tilted white disc with a red spot light shining on it, and believed myself to be seeing a red oval. Today I may believe myself to be imagining the red oval I saw last week, while I am actually imagining the white disc in red lighting I saw. So I am wrong about the shape and color of the thing imagined—without introspection being faulty. A perfectly functioning introspective faculty is simply insufficient to give me knowledge of the color and shape of what I am imagining. Thus, I lack privileged access to even those features of what I am imagining.

Again, the general problem is not that I am imagining a particular, or am engaged in an act of memory. Suppose my environment is littered with tilted white discs in red lighting, and I misperceive them all to be red ovals. If I set out to imagine a red oval of the kind I always see (but no particular one) hanging on the wall of an art gallery (this being a clear case of non-memory imagination), I am still wrong about the actual color and shape properties of the kind of thing imagined.

Granted, this fantastical imagining still involves an element (the white disc) that is “anchored” by a previous act of perception—yet this is the case with the vast majority of fantastical and future-directed imagination (what I will later call ‘modal imagination’). We are almost always imagining things, or types of things, that we have experienced in the past. It does the defender of privileged access no good to back up to cases of imagination that involve no particular or type of thing ever experienced, and to then claim that we have privileged access to *these* imaginative acts since, even if that were true, one would end up having privileged access to only a tiny sliver of what is typically fantastically imagined.

What’s more, there is good independent reason to think we lack privileged access to even this tiny sliver of imagination. For, as I will make explicit in the next chapter, even in the case of imagining kinds of things never before seen, this modal imagining still involves a conditional; it involves representing what certain things (or kinds of things) *would look like if seen*. For instance, the high-jumper represents that the eight foot bar will look a certain way as he approaches and leaps over it. If it would not look (roughly) this way, then he has misrepresented what it would look like—even if he has never seen such a bar before. There is no reason to think that we have privileged access to whether the conditionals implicit in such contents are true; hence, we do not have privileged access to what we are imagining even when imagining wholly new kinds of things.³⁷

4.3 Stepping back – whence the phenomenological intuition?

In lieu of considering further ways of defending the phenomenological intuition, I think we should step back and consider why there is an intuitive pull towards saying that we

³⁷ This idea is further explained and substantiated in Chapters Two and Three, in the sections on modal imagination.

have privileged access to certain aspects of what we are imagining in the first place. If images themselves are the objects of imagination—if they are the things that we imagine—it is easy to see why we might think we have this access. For we typically take ourselves to have privileged access to the facts concerning our own mental states; these are things we can reliably come to know about through introspection alone. And, indeed, common talk of seeing images “with the mind’s eye” makes it tempting to think that imagining consists of introspecting mental images that we create, whereas perception consists of inspecting or becoming aware of external objects.

But why hold that in normal cases of imagination there is a “mind’s eye” that directs itself at a mental image? Perhaps it is because we know the thing represented by the image—the flying cow, say—is not present. Yet few would conclude that when a person hallucinates a flying cow she sees *that* with her mind’s eye as well. A person who hallucinates a flying cow does not see it with her mind’s eye, nor indeed with her physical eyes. This is because to say that *X* is seen implies the presence of an actual *X* (this is why ‘see’ is, in its non-colloquial uses, a success verb). In hallucination I do not see flying cows, I simply visually represent them (from a first person point of view, this may be phenomenologically the same as actually seeing flying cows).

I suggest adopting the same approach to imagination. When I form an image of a flying cow, I do not see the image with my mind’s eye, or with my physical eyes. I simply imaginatively experience a flying cow. On this view, mental images are not the objects of imagination, but are rather what we *use* to imagine generally non-mental entities.³⁸ A similar

³⁸ Colin McGinn stresses this point (2004, p. 64-67): “I no more view my images than I view my concepts, or indeed my percepts...The intentional object of such mental operations as scanning, rotating, and inspecting [images] is the external thing, not the image of it.” McGinn agrees that “we should hold fast to the

point is commonly accepted concerning perceptual experience. Perceptual experiences, it is said, are not the objects of perception. They are not what we are *aware of* in perception.

Rather, they are the mental states in virtue of which we become aware of worldly objects and properties—they are the vehicles or bearers of intentionality, not what the intentional states are of or about. So, I say, it is with imaginative experiences as well.

Why, then, is the “intentional fallacy”—the view that mental features are the intentional objects of perception *and imagination*—so much more tempting in the case of imagination than perception? Perhaps part of it is, as McGinn proposes, due to a “tacit adherence to a sense datum conception of imagining, combined with “the problem of absence” (2004, p. 68-70). Because imagination seems perceptual in nature, and because we only perceive things that are present, it is natural to think that mental images (consisting of sense data) are the present entities that are “perceived” in imagination. Yet, if this were the whole story, we should be equally prone to supposing hallucinatory experiences to be the objects of hallucinatory perceptual experience, since the external objects of hallucinatory experience are equally absent. Yet few contemporary philosophers of mind hold such a position.

I think the principal appeal of the intentional fallacy with regard to imagery can instead be traced to imagination’s distinctive relation to the will. In the case of imagination, we can modify our experience of what is imagined not only by imaginatively exploring it—imagining looking at a sculpture while walking around it, say—but also by “rotating” or

commonsense view that images can be *of* external physical things—so they can be of exactly what percepts can be of...the image is the *vehicle* of intentionality, just as the percept is; it is certainly not the *object* of imagistic intentionality” (2004, p. 45).

“moving” the imagined thing itself, while maintaining a stationary “eye” upon it (as when imagining a rotating cube).

In perception, as in imagination, we can actively explore an object, viewing it from different angles. But we cannot rotate and move the perceived things themselves merely by a mental effort of will (well, the non-telekinetic among us cannot). If we keep our heads and eyes still while perceiving, the only way to change our experience of the object is to reach out and move the object itself (for instance, twirling a cube in one’s hand).

So then, it is a distinctive feature of the *objects of perception* that, if we remain otherwise still, we must move *them* in order to change our experience of them. The thing willfully modulated in such cases is the intentional object. Now, when we imagine objects that are rotating, we do so by willfully modulating our *images*, assuming all the while that there is a stationary “inner eye” viewing the affair. This, I think, leads us to mistakenly conclude (by analogy to perception) that the thing being willfully modulated—the image—must be the intentional object of the act. The image seems to be the thing that is moved around and inspected. But this difference in the relation of the will to our experience is not one that necessitates a difference in the type of things—mental or physical—at which each of perception and imagination is directed. To avoid the intentional fallacy with respect to imagination, we need simply hold that the *imagined object*—a cube—is imaginatively rotated *by means of* modulating an image. The thing imagined, and the thing rotated, is a cube, not an image of a cube.

In light of this, one might justifiably wonder why we do not normally make the distinction, in imagination, between rotating an object while keeping the “mind’s eye” still, and imaginatively walking around (with the “mind’s legs”) a stationary object. Presumably,

the visual phenomenology of each act would be the same in both cases, consisting of a presentation of the same sides of the cube in the same temporal sequence. So it is not that visual phenomenology prevents the distinction. Rather there seems to be a default assumption in the case of imagination that imagining always consists of a stationary “eye” viewing a perhaps changing or moving imagined object. Thus it is assumed that if one’s imaginative experience changes, it is because the imagined object is moving, not an imagined perceiver around an imagined stationary object.

But why are we so much more apt to speak of viewing a rotating image with the mind’s eye than we are to say that we are walking around a stationary imagined object with the “mind’s legs” (not that either way of talking is advisable)? Sure, it’s a matter of custom; to speak of a “mind’s eye” is idiomatic, to speak of “mind’s legs” is not. But what lies behind the idiom? If I were forced to speculate, I’d say that it is grounded in the combination of two views: first, that mental images are the objects of imagination, and, second, that they are instantiated in a two-dimensional visual field, not unlike a television screen. After all, if mental images are two-dimensional, then we can’t get a good view of them from the side; we just have to wait for them to turn in their two-dimensional medium. A healthy set of mind’s legs will do us no good.

4.4 A word on McGinn

Despite sharing my view that perception and imagination have the same kinds of intentional objects (see fn. 17), Colin McGinn strongly disagrees with my claims concerning our epistemic access to what we are imagining. McGinn tells us that “it has been very widely noted that the identity of one’s imaginative object enjoys epistemic privileges” (2004, p. 30).

While he wishes “not to belabor this familiar point,” he does attempt to shed some light on why it is true.

McGinn thinks he has “special access” to the identity of his imagined objects because “the identity of my imagined object is fixed by my imaginative intentions, to which I have special access, while the identity of the perceived object is fixed by what is out there in the external world” (p. 45). In imagination, “the imaginer starts with the object and then constructs an image of it.” This is clear, he thinks, from reflection on his mother: “I know that my image is of my mother because I *intended* it to be; I don’t have to consult the appearance of the person in the image and then infer that I must have formed an image of my mother” (p. 31).

I have already argued that one’s intentions can help fix the identity of an imagined object without one’s having privileged (or “special”³⁹) access to what is being imagined; if introspection cannot tell us which (if any) of various descriptions under which we think of a thing correctly apply to that thing, then introspection will not be sufficient to tell us what we are imagining.

McGinn tries to strengthen his claim that we have special access to what we are imagining by extending the point to the case of thought in general: Note how odd it would be say [*sic*], “You think you have been thinking about your mother but really you have been thinking about someone else”—as if your intention to think about someone played no role in fixing whom you were thinking of. By contrast, my intention to

³⁹ McGinn does not say what “special” epistemic access formally amounts to, or what it is to “enjoy epistemic privileges.” It is safe to conclude from his use of the terms, however, that “special” epistemic access need not be incorrigible, and is more or less the kind of access we have to our own beliefs and desires (see his fn. 41). This agrees closely enough with my notion of privileged access for present purposes.

see my mother could fail utterly—as when someone else steps in front of her. It is up to nature whom I see, but it is up to me whom I think about (or imagine) (fn. 41, pg. 169).

The problem here is that it is *not* very odd to say that a person mistakenly thinks he is thinking about his mother.⁴⁰ Oedipus, when he thinks about his adoptive parents, mistakenly thinks he is thinking about his mother. What *is* odd is to tell someone, “You think you *believe* you are thinking about your mother, but you do not really believe you are thinking about your mother.” This is because people have privileged access to what they believe. But what or who a person is thinking *about* or *of*, just like what a person is seeing or imagining—or, indeed, is believing something *of*⁴¹—occurs at the level of reference, not sense.

Of course, one’s intentions play a *role* in determining both what it is one is thinking of and what it is one is imagining. But, as we have seen, this is not at all sufficient to provide privileged access to what one is imagining (or thinking *of*). While perception and imagination do indeed differ in their relation to the will, this difference does not result in an *epistemic* difference, as so many have supposed.

Section 5: Summing up

Throughout this chapter I’ve given reasons for thinking that imaginative experiences have extra-mental “success” conditions. It is in virtue of such conditions that we can easily be wrong about what we are imagining; for we inevitably lack privileged access to whether those conditions are met. Some of these conditions spring from commitments to how certain

⁴⁰ And one need not, when saying this, be suggesting that intent plays *no* role in fixing who one is thinking of.

⁴¹ As noted earlier, one can believe something *of* a spy, without having privileged access to the fact that one believes that of a spy—when Bill believes that his neighbor James, who happens to be a spy, has nicely trimmed hedges, he believes *of* a spy that he has nicely trimmed hedges, without necessarily believing that a spy has nicely trimmed hedges. Similarly, one can be thinking of a spy while not thinking (i.e. believing) that that person is a spy.

things looked in the past, while other arise out of commitments concerning how certain things would look.

So far, the “success conditions” have been only very generally specified. I offer my own more detailed account in the next chapter. This chapter has aimed to motivate the general view that sensory imagination is at root an assertoric faculty of mind, aimed at the past and the possible, and not a “quasi-perceptual” faculty that invariably involves misrepresenting the present, nor a “fitless” kind of mental state with no real satisfaction conditions.

Finally, to advertise a point that will gain prominence in the next chapter: while it is generally “up to us” to determine what the intentional content of our imaginative experience is—it is up to us what we are *trying* to imagine—it is *not* up to us to determine whether the success conditions given by that content are fulfilled. That is, it is not up to us whether we *succeed* in imagining what we are trying to imagine. The will does indeed play a special role in imagination—it determines the *intentional content* of our imaginative experiences in a way that it cannot in perception. Yet, whether we are in fact imagining the things we are trying to imagine in forming those contentful experiences is, I have argued, another matter.

CHAPTER 2:

The Content of Imaginative Experience

1. The Contents of Imaginative Experience

Having set out in Chapter One various reasons for thinking that imagination has certain general representational features, I now turn to the details of what the representational content of various acts of imagination might be. While what follows is essentially a single proposal for the treatment of imaginative content, there are differences in two kinds of imagination I will identify. The first kind encompasses all imagination involved in pure memory, such as when I imagine my kitchen in order to remember how many cabinets there are. These cases I call ‘memory imagination’. The second kind contains all imagination involved in fantasy or thinking about the future. I call this ‘modal imagination’, and I identify two kinds: hybrid modal imagination and pure modal imagination. To count as pure modal imagination, an imaginative act must be directed at objects one does not believe oneself to have previously seen. Hybrid modal imagination represents objects one believes oneself to have previously seen, situated in novel contexts.

I will first describe how contents are structured for memory imagination, then move to the hybrid modal cases, and from there to pure modal imagination. The discussion of memory imagination is by far the longest because it includes an explanation of basic features of the contents of imagination that are present in all three cases. Before getting to that discussion, however, I’ll first say a few things about the way in which I describe psychological contents throughout this chapter.

On the description of psychological contents

The fact that psychological contents will here be described in natural language does not require or assume that the format of representation of the experiences themselves is itself quasi-linguistic. We can distinguish between the psychological states themselves—states with contents we wish to describe—and our descriptions of those states, which must inevitably be couched in some language (Cf. Searle, 1991, p. 228). My point in describing the contents in natural language is to get clear on the psychological role of the states—it is to reveal what such states in effect “say” to the larger cognitive system, regardless of the underlying cognitive architecture in virtue of which they are able to do so.

So, for all that will be said, forming images and visualizing *may* be a matter of entertaining certain groups of propositions in a discursive “language of thought”, as Zenon Pylyshyn (2002), Daniel Dennett (1981), and other “descriptionalists” have argued.⁴² Or, as Kosslyn (1994), Tye (1991) and Fodor (1975) have urged, imagery may occur in a distinct “pictorial” or “iconic” format.⁴³ Nor have we any reason to think that these options exhaust the possibilities concerning the format in which imaginative experience represents. This dispute concerning the *format* of representation involved in imagery and imagination

⁴² According Pylyshyn (2002), the only thing “distinct” about imagistic thought is that it invariably involves simulating what it would be like to see things. Descriptionalists see no reason to think that this “simulation” could not take place in the same “descriptive” or “discursive” format as all other thought; they also attempt to explain away the empirical evidence that pictorialists bring to bear in support of their view (see, e.g., Pylyshyn, 2002).

⁴³ Spelling out what it means to be a pictorial or iconic representation is not a simple task, but common attributes cited are 1) that, overlooking considerations of fineness of grain, any arbitrarily chosen part of the representation represents a part of thing represented (Fodor, 2003, ch. 2) and 2) such representations invariably represent spatial relations (Tye, 1991), and do so in virtue of preserving those spatial relations in some way in the representation itself—even if, in the representations, the spatial relations are merely preserved in a virtual or “functional” space (Kosslyn, 1994). While many have claimed that that Kosslyn’s notion of a “functional space” is unacceptably obscure (Fodor, 2003; Pylyshyn, 2002), Michael Tye (1991, ch. 3) does an admirable job of explicating Kosslyn’s core idea.

constitutes the traditional “imagery debate” in philosophy and psychology.⁴⁴ That debate is not directly addressed by this chapter—indeed I will do my best to avoid it, until the last Summary and Speculation chapter of this dissertation. For the most part, I am interested in *what* imagery represents, not the format in which it does so.

That said, it is possible that one or another view concerning the format of imagery will be more or less plausible in light of an account of what properties images represent. If that is so, then the present account of what images represent can be seen as constraining theories of image format.

This seems to me a fair way of proceeding—to first determine how mental images function (and, by extension, their representational content), and then move on to the question of what their format might be, given that they have this representational content. For, surely, the question of *how* images represent the things they represent can only be approached given at least a provisional agreement about *what* they represent. This is an accepted order of explanation when it comes to other forms of representation. Supporters of Fodor’s Language of Thought hypothesis infer from the fact that anyone who can think that *X* loves *Y* can also think that *Y* loves *X* that the underlying structure of thought has certain (systematic, combinatorial) syntactic features—this is to move from thought content to thought architecture and format. The same direction of explanation is present in analyses of depictive representation in the arts; the question of *how* a painting manages to represent Napoleon can only be approached given an agreement that it does in fact represent Napoleon.

Nevertheless, constraints might push in both directions—content may well motivate and constrain architecture, yet, once a reasonable account of cognitive architecture is in

⁴⁴ See Block (1981) for a canonical collection of papers on the subject; see Pylyshyn (2002), with responses, for the most sophisticated recent moves on both sides of the debate).

place, it may in turn justify us in revising some aspects of the original account of thought content. It remains a peculiarity of imagery, however, that so much emphasis has gone into investigating its underlying architecture, with so little deep analysis of its representational content. Let us turn, then, to an analysis of *what* imagination represents—to what it “says”.

1.1 Memory Imagination

As we saw in chapter one, both intuition and empirical research support the claim that autobiographical, “episodic” memory⁴⁵ is a form of imagination. Often, one can only remember certain facts about the past (what one had for breakfast, the number of windows in one’s home, etc.) by forming images. Other cases of imagination only implicitly involve pure memory. Take the case where I simply imagine the gardener I met yesterday. I may not explicitly take myself to be trying to *remember* anything about him—perhaps I am just thinking about his peculiar moustache. Nevertheless, if nothing fantastical is added to the scenario, I am implicitly involved in an act of pure memory concerning how a particular person looked to me at some point.

I will use the notation IMG MEM (X) to refer to a memory imagining with the content X, just as the notation BEL (X) is often used to refer to a belief with the content X. IMG MEM (and IMG PM later) is to be interpreted as an *assertoric* attitude, one with a mind-to-world “direction of fit” like belief and perception (if there is a perceptual attitude), and unlike desire. I address below the question of whether the imaginative attitudes I discuss *just are* (or may be reduced to) belief.

⁴⁵ Psychologists typically define episodic memory as “a memory system that supports remembering autobiographical experiences” (Schacter, et al., 2007, p. 657). It is assumed in the psychological literature that this “system” makes essential use of mental imagery (Addis, personal communication).

To see how the contents of the imaginative experiences involved in pure memory should be structured, consider the following example. On Pedro's first trip to Paris, he walks up the Champs Elysees to the Arc de Triomphe, taking time to admire the famous monument. Later in the day Pedro is asked to imagine the Arc de Triomphe, and he attempts to do so. To capture the representational content of his imaginative experience, I suggest the following.

IMG MEM (there was a large white arch with thus and such visible features⁴⁶ AND the Arc de Triomphe looked roughly this way)

This schema mirrors John Searle's theory of perceptual content⁴⁷ (1983, Ch. 2) in having two components: an existentially quantified component ("that there was *a* giant white arch..."), conjoined to a reflexive causal/indexical condition (and I saw the Arc de Triomphe...) that makes reference, via the demonstrative "this way", back to the existential component. If either (or both) of the two conditions is not satisfied, then Pedro is not imagining what he thinks he is. All three forms of imagination I will discuss have a similar two-part structure. I will say a little about each component separately now, and then indicate how they work together.

First, however, it is very important to keep the following distinction clear: while Pedro has wide latitude in determining the representational content of his imaginative act—he decides what he is trying to imagine—it is *not* up to him to determine whether the correctness conditions given by that content are in fact satisfied. He will only be imagining

⁴⁶ As I discuss below, "thus and such visible features" is a placeholder for the properties attributed to an object by the image itself, as opposed to the additional properties specified by the second condition (which serve to determine which *particular* object one is trying to imagine).

⁴⁷ On Searle's theory, perceptual content possesses an existential component, "that there is *an* X at Y," conjoined to an indexical reference the present experience and its cause: "that there is an X at Y is causing this visual experience." Searle's theory is often criticized as being too complex or cognitively sophisticated (Burge, 1991); the present account of imagination may raise similar worries. I will wait to address these concerns in detail in Section 6 when the theory has been more fully developed. Only some of the worries that arise for Searle's account ultimately pertain to this one.

what he is trying to imagine (or what he believes he is imagining) when those conditions are in fact satisfied. In this way, we respect the idea that the *will* is involved in imagination—it helps supply the intentional content of an imaginative experience—while also allowing for the possibility that what one is trying to imagine is not in fact imagined (because the mind-independent success conditions are not met). This key facet of the account will become clearer as we proceed.

1.1.1 The Existential Component as “the image”

We can now take a closer look at the two components of imaginative contents . Looking at the existential component first (in the example above), notice that it involves a past-tense verb—‘was’ instead of ‘is’. This is because, intuitively, Pedro can successfully imagine the Arc de Triomphe even if it was destroyed after he saw it. If ‘is’ were part of the existential component, and if no arch of the kind Pedro is trying to imagine exists when he’s trying to imagine it, then the experience’s correctness conditions would not be satisfied. We would have to say that he failed to imagine the Arc de Triomphe. Intuitively, for Pedro to imagine the Arc de Triomphe, it is not necessary that there *is* any such arch—only that there *was* one.⁴⁸

The ‘was’ also respects the fact that, in imagining the Arc de Triomphe, Pedro need not be pretending that he (or someone else) *is* currently seeing the Arc de Triomphe. Thus it is consistent with the fact that he is not disposed to act as though the Arc de Triomphe is

⁴⁸ Currie and Ravenscroft miss this point when they suggest that “if I have a visual image of a mountain, then the content of my imagining is...there being a mountain somewhere” (2002, p. 26). This view implies that if there is *not* a mountain “somewhere”, then the imaginative experience misrepresents, or is nonveridical. But surely all mountains could crumble to the sea—and one could even *believe* them to have so crumbled—without this impacting one’s ability to imagine a mountain. It is one thing to believe (or perceive) that there is a mountain somewhere, quite another to imagine a mountain.

present, while allowing us to avoid holding that imagination always involves non-veridical representations that are in some sense cognitively “resisted”.

Notice also that the existential content does not specify *which* large white arch it is that is being imagined. This allows the existential component, in imagination, to play the role we pre-theoretically assign to the “image”. Suppose Pedro later sees another arch qualitatively identical to the Arc de Triomphe—say, a perfect duplicate of it at Disney’s Epcot World Center. When he goes to imagine this arch, he may use what intuitively seems like a token of the same image type that he used to imagine the Arc de Triomphe. This *prima facie* datum can be explained by the fact that the existential component of his imaginative experience will be the same in both cases—namely, “that there was a giant white arch with thus and such visible characteristics.”⁴⁹ This point requires further discussion, particularly with regard to the issue of the often alleged indeterminacy of mental images.

1.1.2 Digression on the indeterminacy of mental images

This account of the content of the imagistic portion of an imaginative experience allows for one kind of indeterminacy in the image, but not another perhaps more pernicious kind. Let me explain. There is a principled limit to the kinds of properties that will be represented by the images of any given individual: whatever properties of an object a person cannot visually discriminate will not be represented by the image he uses to imagine that object (this is not to say, however, that all of the properties a subject *can* visually

⁴⁹ Of course, to account for all of the visible properties represented by an image, the existential component of the above contents would normally need to be filled out significantly. Not only is a large, white, arch-looking thing represented by Pedro’s image, but certain sides of it are represented, and it is cast in a certain pattern of shadow, etc. Nevertheless, all of this perspectival detail can, in principle, be packed into the existential portion of the content. What will *not* be part of the existential content are properties not available to vision (e.g. the molecular structure of the Arc de Triomphe).

discriminate are represented by the image he uses to imagine the object⁵⁰). An image that represents a giant white arch will normally be indeterminate for properties such as the arch's date of creation, or its height in centimeters. This indeterminacy extends to the *particularity* of the object as well; since there can almost always be another object that is visually indiscriminable from any particular object *X*, the particularity of the object—which token object it in fact is—will not usually be part of the content of the image used to imagine that object. As already noted, this is accommodated by the fact that it is an existential content, declaring the past (or, I will suggest, possible) existence of *a* or *some X* with thus and such properties that forms the imagistic portion of imaginative content.

These kinds of indeterminacy, however, are not unusual. Sentences in natural language can be indeterminate in the same way. 'A dog jumped over a fence' does not go into great detail concerning the dog (e.g. its breed) and fence (e.g. its color), nor does it specify *which* dog jumped over which fence. Yet there is a more radical indeterminacy sometimes alleged of mental images, whereby tokens of the very same image-type can represent completely different properties or states of affairs. Note again the distinction I appealed to earlier: there is the question of what the image represents, on the one hand, and the question of what one's overall imaginative experience represents, on the other. If one fails to appreciate this distinction, then the (intuitive) presence of the same image in acts of imagining different things may lead one to suppose that the image itself somehow leaves radically indeterminate what is represented.⁵¹

⁵⁰ The matter of which properties are represented "directly" by vision and which are merely *inferred* from the way a thing looks is complex and controversial. I will not try to settle the issue here. See Siegel (2007) and Prinz (2007) for unusually expansive accounts of the kinds of properties represented in perception.

⁵¹ This mistake can occur even if one does have that distinction in place, if one still holds that *all* of the intentional properties relevant to one's imaginative experience are contained in the image itself. For instance, Paul Noordhof holds that "a mental image may have intentional properties that are not settled by the character

For instance, in a famous passage Wittgenstein writes: “I see a picture, it represents an old man walking up a steep path leaning on a stick.—How? Might it not have looked just the same if he had been sliding downhill in that position?” (1953, *ff.* 137, fn. 2(b)). Later he asks, “What makes my image of him into an image of *him*? Not its looking like him” (1953, p. 151). Many have taken such musings to suggest that mental images cannot be the bearers of truth (or reference), since a single type of image can seemingly serve too many different representational purposes (Fodor, 1975, Ch. 8). On this view, images function more like symbols than thoughts; they possess their intentional properties only derivatively, through an act of interpretation or stipulation (or through convention).⁵²

Though attractive, this view raises a host of problems. As an *a priori* matter, there is clear tension in arguing for the representational ambiguity of mental images from the ability of multiple tokens of a particular image type to represent multiple objects or properties. For if tokens of the same image type really can represent different things (and therefore possess different intentional properties), it is unclear why they should have been considered images of the same type to begin with. And if they are not the same image, there is no reason to suspect that they are in some way referentially or semantically ambiguous. One could reply that images are typed only by their non-intentional properties, and that this is how tokens of the same image type can be featured in different representational tasks. Yet, in addition to

of the image alone” (2002, p. 431). This is a questionable way of putting things, since an image plausibly changes type as soon as it acquires new intentional properties, making it misleading at best to speak of an image having intentional properties “not settled by the character of the image alone.”

⁵² Fodor (1975, Ch. 8) argues that images are generated *out of* descriptions—i.e., that there are cognitive mechanisms that take descriptions as inputs and give, as outputs, images that represent what is represented by the descriptions. For the reasons I discuss in the text (relating to imagery’s necessary role in memory and planning) this cannot be the whole story; there must be some mental images that possess their intentional properties intrinsically (as Fodor admits may be the case (1975, p.193)). For if all information represented by images is first available in a quasi-linguistic descriptive form, the actual formation of imagery would be inessential to most memory and planning tasks. Fodor’s argument that forming images allows the cognitive system to run multiple tasks in parallel gives no explanation for why imagery would be *required* for some representational tasks (e.g. memory), and not merely expedient.

clashing with the strong intuition that images are always *of* things (and therefore bear intentionality essentially), an alternate account of how we ought to type images would then be required. This would need to be an account that showed why imagery is crucial to reasoning and memory if it does not in itself possess intentional properties.

More pressingly, if mental images do not possess their intentional properties intrinsically—if we must interpret them in order for them to mean what they do—then why, as Rubin and Greenberg found, is forming them necessary (as opposed to a merely expedient) for remembering one's own past? And why, as the work of Schacter *et al.* (2007) and Addis *et al.* (2007) suggests, is imagistic thought essential to one's ability to think about one's own future?

Fortunately, these questions can be avoided—and the Wittgensteinian intuition accounted for⁵³—if we simply make use of the image-content/imaginative experience content distinction. This allows us to make sense of the intuition that mental images can serve different representational purposes, without falling into the view that mental images are symbol-like entities, requiring interpretation, and without concluding that images cannot be typed by their intentional properties.

1.1.3 The causal/indexical condition, and its relation to the existential component

Of course, if images are always of *an* X or Y, and not of *the* X or Y, then we are stuck with the counterintuitive result that we never imagine particulars. This problem is solved by

⁵³ If a picture of a man on a hill could truly look the same whether the man is climbing or sliding, then perhaps the matter of his climbing or sliding is not represented by the picture (maybe it's just a picture of *a man leaning forward on a hill*). As for what makes his image of a certain person of *that person* (and not someone who merely looks like that person)—perhaps the image itself is not of a particular person (it is rather of *a person* with certain features), and his imaginative act as a whole is of a certain person.

the reflexive causal/indexical component. In the example above, “The Arc de Triomphe looked roughly this way,” explicitly identifies the particular object that the imaginer is trying to imagine. Differences in the causal/indexical component will determine whether the context of the imaginative experience is one of imagining the Arc de Triomphe, the Epcot Arch, or some other arch. In this way, the *imaginative experience as a whole* can be directed at a particular, even if the existential “imagistic” component does not specify the particular. This means that, properly speaking, we should not say that Pedro’s image in the case above is an image of the Arc de Triomphe. Rather, it is an image of a large white arch with thus and such other visible properties, while *his imaginative experience as a whole* constitutes an act of imagining the Arc de Triomphe (so long as the two conditions are indeed met).⁵⁴

Looking again at the causal/indexical reflexive condition—“I saw the Arc de Triomphe looking roughly this way”—the demonstrative phrase “this way” refers to the way the existential component of the experience represents a thing as having looked. The “roughly” that qualifies “this way” is needed because, intuitively, when we imagine a thing, our experience need not recreate (or simulate) our experience of the thing in precise detail. It need only provide an approximation of what it was like (or what it would be like) to see the

⁵⁴ Others have suggested a similar interplay between “imagistic” and non-imagistic representations in imagination. Martin (2002, p. 403) notes, “The same imagery can be put to different imaginative purposes: one might imagine red apples, perfect wax replicas of apples, the skins of such apples with the cores hollowed out or a cunning illusion of the presence of apples, while visualizing in the same way. The differences between these cases lies not in the sensory core of imagining but the way in which that core is used in make-believe, the way in which it has, so to speak, been labeled.” Christopher Peacocke offers a similar distinction between the content of images themselves and what is “S-imagined”. While S-imagining “is not literally supposing, it shares with supposition the property that what is S-imagined is not determined by the subject’s images, his imagined experiences,” (1985, p. 25). Neither Peacocke nor Martin, however, develop these suggestions in a way that would enable us to properly understand the satisfaction conditions of imagination.

thing. It is not the role of imagination to recreate the exact sort of experience one has when seeing.⁵⁵

Finally, the verb “saw” is important to memory imagination in indicating that a particular act of visual perception is believed to have occurred. For a person to see something, her visual experience of that thing must be caused in the right way. I will not try to spell out what the “right way” is here—that is a matter for a separate analysis (see, e.g. Lewis (1980)). I will assume, however, that seeing a picture of a thing (where the picture is of that thing because it is caused in the right way...) counts as an instance of seeing the thing pictured. This much falls out of common usage; it seems odd to say that a person who has studied photographs of the Eiffel Tower has never seen the Eiffel Tower (though it is less odd, interestingly, to say that one who has only seen drawings of the Eiffel Tower has never seen the Eiffel Tower). But, again, nothing peculiar to this account hangs on this phenomenon. If one denies that seeing a photograph of a thing counts as actually seeing that thing, then one will simply deny that one can imaginatively remember things one has only seen pictures of (it will be, instead, the pictures themselves one is imagining/remembering).

We can now draw an explicit connection between these contents and the intuitive claim that people can be wrong about what they are imagining. When the causal and existential conditions are both fulfilled, it is the thing (or kind of thing) specified by the causal condition that is imagined. In the case above, both conditions are met, so we say that it is the Arc de Triomphe that Pedro imagined.

⁵⁵ Imagery is often thought of as a weak or degraded kind of vision, due to its relative lack of informational detail. Current research on the role of imagery in memory and problems solving (Schacter *et al.*, 2007; Kosslyn, 1994), however, suggests different point of emphasis: we can view imagery as a rarified kind of vision, one which abstracts away from unnecessary detail and which can therefore be put to more diverse uses with less cognitive effort. See Arp (2008) for an account of imagery’s evolutionary history in humans that fits well with this “rarified” conception.

If one or both of the conditions are not met, we say that the thing specified by the second condition was *not* imagined (it was only *believed to have been* imagined). This occurs in the following sort of case. Suppose that Roger has heard of, but never seen, the Arc de Triomphe. One day he is put under sedation and awakened at the Epcot World Center's Paris Pavilion, which features an arch that looks just like the Arc de Triomphe. A man in a striped shirt and beret tells him that he is actually in Paris, viewing the Arc de Triomphe. Disoriented, Roger believes him. He is then sedated again and taken back home to New York. The next day Roger is asked to imagine the Arc de Triomphe, and tries to do so. The content of his imaginative experience is the same as Pablo's was above, namely:

IMG MEM (there was a large white arch with thus and such visible properties and the Arc de Triomphe looked roughly this way).

In this instance, however, Roger *fails* to imagine the Arc de Triomphe—he believes himself to be imagining something he is not—because the second component of the content is not satisfied. It is not true that he saw the Arc de Triomphe to look the way his image represents an arch as looking. Rather, he saw the Epcot Arch at the Paris Pavilion to look the way that his image represents an arch as having looked. It is therefore the Epcot arch that he is truly imagining.⁵⁶ And, indeed, were we to reveal to Roger what had happened to him, he would likely agree that he had been imagining the Epcot Arch all along.

⁵⁶ If we suppose that Roger has seen a variety of arches that look the way his image represents an arch as looking, but *not* the arch referred to in the causal condition of his imaginative act, it may be indeterminate *which* arch (if any) he is in fact imagining. The issue can potentially be settled by looking at the other descriptions (if any) under which Roger thinks of the arch he is trying to imagine (we could interview Roger to figure this out). If these descriptions unambiguously pick out another arch that he has in fact seen, it is plausible to say that this is the arch that he was in fact imagining.

It is important to keep the following distinction clear: while it is up to Roger to determine the representational content of his imaginative act—he decides what he is trying to imagine—it is *not* up to him whether the correctness conditions given by that content are satisfied. Thus, it is not up to him whether he really is imagining what he is trying to imagine and may believe he is imagining. If Roger had merely tried to imagine the large white arch he saw yesterday, and remained agnostic over whether he was imagining the Arc de Triomphe, he would have succeeded. In such a case, the content would look like this:

IMG MEM (that there was a large white arch with thus and such visible properties and a giant white arch looked roughly this way yesterday).

Both of these conditions are satisfied and, intuitively, it seems right to say that Roger imagines the arch he saw yesterday. Thus, the intent of the imaginer can change the correctness conditions of an imaginative act, even in cases where the imagistic element of the act (which contributes to the existential component of the content) remains the same.

The key role that the imaginer's intent plays in fixing the satisfaction conditions of an imaginative experience helps explain the intuitive sense in which the nature of an imaginative project is "up to the imaginer". It is up to Roger to stipulate whether it is the Arc de Triomphe, or simply *an* arch that he saw, and which he is now trying to imagine. It is also up to him to determine which existential content he calls to mind. What he *does not* get to decide is whether his imaginative act as a whole succeeds in hitting its target. He does not get to stipulate that the correctness conditions are indeed satisfied.

1.2 Modal Imagination -- Hybrid

Very often when we imagine things, we are imagining objects or kinds of things that we have seen in the past, but in some new setting, and perhaps with slightly different visible properties. These constitute cases of Hybrid Modal (“HM”) Imagination. In HM imaginings, we do not set out to imagine wholly new objects or scenarios, but rather “reshuffle” old things in new ways.⁵⁷

These imaginative experiences will naturally have correctness conditions that are partly tied to past visual experiences. As an example, consider again Pablo, who earlier succeeded in imagining the Arc de Triomphe. Were Pablo not simply to try to imagine the Arc de Triomphe, but, more ambitiously, to try to imagine the Arc de Triomphe painted red (preventing this from being a case of memory imagination), the HM content would be as follows:

IMG HM (there *could be* a giant, red arch with thus and such visible features AND the Arc de Triomphe and the color red looked roughly this way).

Here the first half of the content contains a modal notion—that there *could be* an arch of a certain kind—instead of the past tense verb ‘was’ that is featured in acts of pure memory imagination. This respects that fact that Pablo is here not committed to the Arc de Triomphe having been or actually being red, but only to the idea that it could be red. The

⁵⁷ Of course, an empiricist of the Humean variety might contend that *all* modal imagination is of this kind, consisting of “ideas” (faint copies of sensory “impressions”) recombined in various ways. In positing a category of pure modal imagination, I have allowed that not all imagination need involve mental states that achieve their meaning (or reference) through a causal link to prior acts of perception. This is because, intuitively, I do not see any reason why one *must* always be committed to one’s imaginative experience “matching” or “deriving from” some prior experience.

second condition, however, is very similar to what we find in the pure memory cases; it picks out one, or several, objects or properties that the subject takes himself to have seen, and which determine the particular objects and properties that the imaginer is trying to imagine. Supposing that Pablo is correct about the relevant past perceptions, and that there could be a red arch of the kind represented by his image, then Pablo will in fact imagine the Arc de Triomphe painted red. If, however, Pablo, like Roger, were mistaken about the nature of his past perceptions and had only ever seen the Epcot Arch⁵⁸, or had not in fact seen red things, then he would not succeed, in this case, in imagining the Arc de Triomphe painted red.

Again, the claim that Pablo's imaginative acts fails is supported by considering what Pablo would say about his imaginative act, were the true nature of his past perceptual history revealed to him. Were it revealed to him that, contrary to his beliefs, he had never seen the Arc de Triomphe, but had only seen the Epcot Arch, then, intuitively, he would conclude that it was the Epcot Arch he imagined painted red and not the Arc de Triomphe.

Yet, what if Pablo has not seen the Arc de Triomphe—has only seen the Epcot Arch—and knows all this full well? Can he not still imagine the Arc de Triomphe painted red? Intuitively, it seems he can. Here I suggest that his beliefs about his past experiences change the nature of his imaginative experience. He does not take himself to be *remembering* what something actually looked like; rather, he is imagining what it would be

⁵⁸ Here I am assuming that seeing the Epcot arch is not equivalent to seeing a picture of the Arc de Triomphe, even if it was designed to be a copy of the Arc de Triomphe (and so is, in a sense, caused by the Arc de Triomphe). This brings up a question: what if someone has seen a (life size) picture of the Arc de Triomphe but believed himself to have been seeing the actual Arc de Triomphe, and then tries to imagine the Arc de Triomphe (and he had never before seen the Arc de Triomphe). Is he unwittingly imagining the picture, or successfully imagining the Arc de Triomphe (since, normally, we are assuming that seeing a properly-caused picture of a thing allows one to imagine the thing itself)? Here it seems intuition could lead in either direction, though I am inclined to interpret it as a case of successfully imagining the Arc de Triomphe.

like to see some thing he has never before seen. The distinctive intention behind his act turns it into a case of pure modal (“PM”) imagination, to be discussed next.

1.3 Modal Imagination -- Pure

So far I have considered cases of imagination that are memories and that are hybrids of memory and fantasy. The next step is to consider imagination that involves imagining objects one has never seen. Let us assume, for the sake of argument (and as intuition would have it), that we can indeed imagine things never before seen. We can return to Pablo, this time supposing the he has never seen the Arc de Triomphe, but wishes to imagine it. Let us suppose also that he knows certain relevant things about how it looks—e.g., that it is large, white, and arched. The content of his imagining could be along the lines of:

IMG PM (there could be a large white arch with thus and such visible properties AND the Arc de Triomphe would look roughly like this)

The difference between PM imagination and HM imagination is that, in PM imagination, *both* conditions contain a modal. Not only does the existential content contain the ‘could’ modal, but the causal reflexive condition also contains the modal ‘would’: “the Arc de Triomphe *would* look roughly like this” (where ‘this’ refers to the way the existential content represents a possible *X* as looking). Once again, the thing that one is attempting to imagine will be the thing referred to in the causal/reflexive condition, and one will succeed in imagining that thing if both conditions are satisfied.

As with memory and HM imagining, the success conditions for PM imagination can fail to be met. One or both of the counterfactual statements might be untrue. If there could be no such arch, or if the Arc de Triomphe would not look roughly the way the image indicates it would look, then Pablo is not imagining the Arc de Triomphe.

To see how modal imagination can go wrong, suppose that Pablo thinks that the Arc de Triomphe is a thin, shiny, silver arch not unlike the Gateway Arch in St. Louis. The content of his imaginative experience would be:

IMG PM (there could be a thin, shiny, silver arch with thus and such other visible properties
AND the Arc de Triomphe would look roughly like this)

Here it seems the second condition would not be satisfied—the Arc de Triomphe would not (absent major modifications) look roughly the way the existential portion of the content represents a possible arch as looking. Pablo has not imagined the Arc de Triomphe after all.

However, one may object that there is another way of reading this situation. One might argue that Pablo does indeed here imagine the Arc de Triomphe, he simply does not imagine it as it actually looks. Rather, he imagines it as looking roughly like the Gateway Arch. This may even seem supported by the fact that, were Pablo to later see the Arc de Triomphe, he may say, “It is not as I imagined it”—for this implies that he did in fact imagine *it* (and not some other arch) before.

To settle this issue, we need simply distinguish between two different possible imaginative projects. In the first, Pablo is trying to imagine how the Arc de Triomphe would look, were he to see it as it actually is. In another, he is trying to imagine how the Arc de

Triomphe would look were he to see it having the perceptible features of the Gateway Arch. The content of the first project is outlined above. The second would be as follows:

IMG PM (there could be a thin, shiny, silver arch with thus and such other visible properties AND the Arc de Triomphe, made to have the features of the Gateway Arch, would look roughly like this)

Both imaginative projects have clear—and different—correctness conditions. In the former case, the imager is trying to represent how something looks were he to see it *as it now is* (this is the default mode of imagining). In the latter the imager is trying to represent how something looks were it to have certain features that it currently may or may not now have—this is a more unusual imaginative project, and involves a slightly more complex content that takes it beyond the “default” case.

So, depending on the context and the purposes of the imager, some instances of a person saying “it is not as I imagined it” will arise from situations where certain correctness conditions were not fulfilled (as in the first instance⁵⁹), while others will arise from situations where a person purposefully imagines an object to have perceptible features it does not in fact have. These latter cases will be veridical cases of modal imagination, so long as the object mentioned in the second condition *really could* have such features.

Of course, it is arguable that the Arc de Triomphe could not possibly remain the same entity while taking on the perceptible features of the Gateway Arch. If that is true, then even a person who purposefully sets out to imagine the Arc de Triomphe having the superficial

⁵⁹ In which cases, strictly speaking, it will be false that one imagined what one was trying to imagine (since the second condition is not fulfilled).

features of the Gateway Arch will fail to imagine the Arc de Triomphe (whereas, intuitively, there is no comparable problem in imagining the Arc de Triomphe painted red). The upshot is that, on my account, the impossible is *unimaginable*—even if we can *seem* to imagine impossibilities. Obviously this raises a whole host of questions concerning the relation between possibility and imaginability, and concerning modal epistemology more generally.⁶⁰ I deal with these issues in detail in Chapter Three, using as a point of comparison Kripke's (1980) account of *a posteriori* necessities and the modal epistemology sometimes thought to underlie that work.

For now I will set those questions aside in favor of clarifying and defending other aspects of my account.

2. A Comparison to Peacocke's Proposal

It may help to elucidate my proposal by comparing it to a well known rival account in the literature. Peacocke (1985), citing Wittgenstein (1969), argues that there is a sense of 'imagining' according to which it is absurd to claim that a person is wrong about what he is imagining. However, part of making that argument, Peacocke claims, involves distinguishing this sense of 'imagining' from another one, of which it is not absurd to make this claim. We can ask what a mental image is causally *of*, Peacocke observes, in the sense that a certain actual scene is responsible for the layout of properties represented by an image, due to a past act of perception. It is easy to be wrong about what one's image is of in *that* sense. Writes Peacocke:

⁶⁰ Among these are the questions of whether we can coherently entertain metaphysically impossible propositions as parts of these contents, and what we should say if the conditional in the second half of the content is "vacuously true" (because its antecedent is false). I address these and related issues in chapter three.

The tourist to whom New College is introduced as Hertford College may well, when asked to imagine Hertford College in the winter, form an image which is causally of New College (1985, p. 27).

In this case, the tourist will be wrong about what his image is (“causally”) an image of. *If* one is imagining whatever one’s images are causally of, then one can easily be wrong about what one is imagining. Peacocke distinguishes this sense of ‘imagining’ from another, which he calls “S-imagining” (where “S” is for suppose”). It is absurd, on his view, to ask whether one is correct about what one is S-imagining.

I will have more to say about S-imagining shortly; first let me take up the issue of what images are causally *of*, and its relation to our epistemic access to what is imagined. The way of being wrong about what is imagined described in the above quote may seem similar to the examples I have given. However, there are crucial differences. Note that while I do hold that a past causal interaction is relevant to what one is in fact imagining *in certain (memory and hybrid modal) instances*, I do not hold that what we are imagining are always the concrete particulars from which our images causally derive, and that *this* is why we can be wrong about what we are imagining. I argue instead that we are wrong about what we are imagining when we believe ourselves to have seen a particular thing to look a certain way, yet did not see that thing to look that way (IMG MEM and IMG HM), or when we represent that a thing would look a certain way if it were seen and it would not look that way (IMG MOD). This allows us to have privileged access to the representational content of our imaginative experiences (images included) even if we are sometimes wrong about what we are imagining.

This difference is important for the following reason. On the way Peacocke understands being wrong about the objects of imagination, being wrong does not consist in misrepresenting anything. This is because the representational content of an image, if images are to serve in psychological explanations, will not include many properties of the objects from which the images causally derive. For instance, an image that is causally of the Empire State Building will also be an image of Pablo's favorite skyscraper. Thus, anyone who forms an image that is causally of the Empire State Building will be forming an image that is causally of Pablo's favorite skyscraper. But it is implausible to suppose that one's image represents the building as being Pablo's favorite skyscraper. So, merely knowing what an image is causally of does not tell us what the image represents—what its intentional content is—for the person entertaining it. This is especially clear when we consider images that causally derive from hallucinatory or illusory experiences. If Jim misperceives a blue wall to be red, the image that causally derives from that experience will be one that represents a red wall, not a blue one.⁶¹ Thus, one can be wrong about what one's image is causally of—one can deny that it is of Pablo's favorite skyscraper, or of a blue wall seen to be red—without *misrepresenting* anything while entertaining that image. If being wrong about what you are imagining does not involve misrepresentation, then this sort of corrigibility is a relatively uninteresting phenomenon from the perspective of intentional psychology.⁶²

My account differs in allowing for both privileged access to the representational content of imaginative experiences for subjects, and failure of the satisfaction conditions

⁶¹ Insisting to the contrary that the image nevertheless represents a blue wall has the price of distancing the subject from the representational content of the image. If subjects lack privileged epistemic access to the representational content of their images, then we cannot plausibly appeal to the use of such images in reasoning and memory, etc.

⁶² Nor does Peacocke claim to have identified a surprising or significant feature of images.

given by this content without the subject's knowledge. So, despite the superficial similarities in the cases described by Peacocke and myself, the interpretation we give of them is very different: Peacocke points to a relatively uninteresting feature of our epistemic relation to the historical causes of our images, whereas I use the cases as a springboard for understanding the intentional content of imaginative experience.

Now to consider Peacocke's conception of S-imagining—is it similar to anything I have proposed? To understand S-imagination, we first must take stock of another aspect of Peacocke's view. Peacocke observes that the “same conscious, subjective image” will serve to support numerous imaginative projects, such as:

Imagining being at the helm of a yacht; imagining from the inside an experience as of being at the helm of a yacht; and imagining from the inside what it would be like if a brain surgeon were causing you to have an experience as of being at the helm of a yacht (p. 19).

His explanation for this is that all these projects involve imagining scenarios where the same kind of experience is undergone (recall, I simply appeal to identical existential contents—contents that need only represent worldly properties had by *a* thus-and-such). He moves from there to the claim that all sensory imagining involves imagining experiences. That controversial claim aside, we can now assess his notion of S-imagining. “‘S’ is for ‘suppose’,” Peacocke tells us. “Although S-imagining is not literally supposing, it shares with supposing the property that what is S-imagined is not determined by the subject's images, his imagined experiences” (1985, p. 25). He calls on S-imagining to explain “the difference between imaginings which, though having a common image, still differ” (p. 25). The relevant differences, on his view, “result from different conditions which the imagined

experience is imagined to fulfill,” and these differences are “differences in which conditions are *S-imagined* to hold.”

So, what we S-imagine is a matter of the conditions we suppose to be true of the imagined experience and its setting. The conditions that are S-imagined are not limited merely to potentially different causes of the imagined experience (e.g., was it caused by an evil demon? is it a veridical perception), but to other features of what is imagined as well. To use another of his examples, the difference between imagining a suitcase and imagining a cat wholly obscured by a suitcase will still be a difference in what is S-imagined, even though the S-imagined *causes* of the respective imagined experiences will potentially be the same. So, what is imagined *in toto* will be the imagined experience plus whatever additional S-imagined conditions are stipulated.

It is absurd, according to Peacocke, to think that one might be wrong about what one is S-imagining—this is how he accounts for the intuition that we are always right about what we are imagining. How does this relate to my view? There is one way in which we could be seen to agree. I allow that we have privileged access to the representational content of our imaginative experiences; knowing this could be construed as knowing what we suppose (or believe) ourselves to be imagining, from case to case. Consider the person having an imaginative experience:

IMG MEM (there was a large white arch with thus and such visible properties AND the Arc de Triomphe looked roughly this way).

The person having this experience believes himself to be imagining the Arc de Triomphe, and forms this belief because he has privileged access to the content of his imagination. On my view, it will be absurd (or peculiar, at least) to ask him whether he is sure he believes he is imagining the Arc de Triomphe. Now, if what one is S-imagining is just a matter of what one believes or supposes one is imagining, then I agree with Peacocke that it is absurd to think someone might be wrong about what he is S-imagining (in the same way it would be absurd to ask someone whether she is wrong about what she seems to be seeing).

But there can be no completely clean comparison to my view, however, since Peacocke does not allow a distinction between what one is *actually* imagining and what one is S-imagining, whereas I do. So if we agree about subjects' epistemic access to what they S-imagine, this is only because we disagree about whether there is a further distinction to be made between what is S-imagined and what is actually imagined. On Peacocke's view, what one is actually imagining is fully determined by representational the content of one's S-imagining combined with the intentional content of the image used to imagine a particular experience—these arrows always hit their mark, there is no possibility that they might misrepresent. I have already aired (Chapter 1, Section 3) my grievances with this sort of picture.

By contrast, I hold that two individuals can have imaginative experiences with the same content, while only one of them is imagining what he believes himself to be imagining (see the example above in Section 1.1)—just as two individuals could have perceptual experiences with the same content, while only one of them is veridically perceiving.

As for the Peacocke's suitcase example, I hold that the two imaginative acts (of imagining a suitcase, and of imagining a suitcase obscuring a cat) have distinct contents, but

share the same existential “imagistic” portion of the content. Thus, the first may be written as:

IMG PM (there could be a suitcase with thus and such features AND a suitcase would look roughly this way).

The second would differ as follows:

IMG PM (there could be a suitcase with thus and such features AND a suitcase obscuring a cat would look roughly this way).

3. Complexity Concerns

Some may feel that, as with Searle’s account of perceptual content, the posited contents of imagination are implausibly complex or sophisticated. In the case of perception, a popular simplifying move is to remove from the content the self-reflexive correctness condition that, in Searle’s case, refers to “this experience”. While one’s perceptual experience does need to be caused by in the right way for it to be veridical, it has been suggested that this condition can operate on the sidelines, so to speak, as a kind of governing condition on all perception, and not as part of the content itself as Searle supposes (Burge, 1991). Part of the justification for this move is that the causal condition itself need not be invoked to explain my reasoning and behavior in the vast majority of cases. To understand how a visual experience causes me to reach for a ball, say, it is usually enough to attribute to me the representation *that there is a ball there*, not that I think that the ball is *causing* my current visual experience.

Such a move is not available in the case of imagination, however, since imagination relies upon the will in a way that perception does not. The imaginer’s ability to pick what he will try to imagine creates a large variety of possible conditions of satisfaction; unlike

perception, there is no single “default” background condition that must be met each time. In forming an image of a large white arch, I might be trying to imagine the Arc de Triomphe, the Epcot Paris Arch, or some future Arch that will be constructed in Milan. Unless I stipulate, as part of the intentional content, which it is that I am trying to imagine, the correctness conditions for the act cannot be settled. Indeed, not only can we not determine the correctness conditions, but we cannot anticipate what role the imaginative act will play in reasoning. Again, this stipulation is not needed in perception, as the perceiver’s intent plays no role in shaping the contents of his perceptual state, nor in how that state may influence behavior. Because determining the accuracy conditions of imaginative experience requires taking intent into account, the contents of imaginative experience will naturally be more complex than those of perception.

Nevertheless, the complexity of Searlian perceptual content raises doubts not only because a causal condition does not seem needed to explain everyday human behavior, but because less cognitively sophisticated organisms—organisms likely lacking the concept VISUAL EXPERIENCE that is featured in Searle’s self-reflexive perceptual contents (see fn. 24)—seem quite capable of perception. One response to such a worry is that perception (and imagination) has nonconceptual content—i.e. correctness conditions whose description involves concepts not necessarily possessed by the experiencer (Bermudez, 2008). Defenders of non-conceptual content will simply hold that organisms *can* entertain contents the specification of which involves conceptual capacities beyond their reach, and that we should therefore not find the conceptual complexity of those contents troubling.

This response is similar to aspects of Searle’s defense of his account of perceptual content:

I am not claiming that the perceiver thinks to himself in words the sentence in the brackets or even that the perceiver has any consciousness of this articulation of these conditions at all.

The theoretical representation is a second-order characterization of a set of first order psychological facts. I am trying to specify *in a language* the content of a *nonlinguistic visual phenomenon*; for this reason the complexity is a matter of the theoretical representation and not a matter of the first order psychological facts (1991, p. 228).

The important distinction Searle makes is between “first order psychological facts”—the perceptual and imaginative experiences we are theorizing about—and our “second-order” theoretical characterizations of those facts. According to Searle, by using a complex English sentence to describe the content (and thus the functional role) of a mental state, we do not imply that the state plays that role in virtue of possessing quasi-linguistic properties. Just as one can provide a second-order theoretical description (in English) of the functioning of a spark plug, without implying that the spark plug plays that role in virtue of deploying the concepts featured in that description, so too, suggests Searle, can one use an English sentence to describe the content of a mental state without concluding that this content is grounded in the organism’s possession of the concepts mentioned in this sentence.

Put in these terms, Searle looks to be an early advocate of non-conceptual content. Of course, the doctrine of non-conceptual content faces some pressing questions. Most pointedly, one might wonder how non-conceptual imaginative and perceptual states would manage to interact with the conceptual states involved in reasoning to influence—and *justify*—an organism’s behavior (McDowell, 1994). Defenders of non-conceptual content have developed various lines of response to this criticism (see, e.g., Heck, 2000). To the degree that a plausible story can indeed be told that makes sense of the interaction between

conceptual and non-conceptual states, the complexity objection to the present account becomes less pressing.

Yet, even if one is hostile to the idea of nonconceptual content, the complexity worry is far less pressing in the case of imagination than for perception for at least two reasons. First, the present account of imagination does not require imaginers to deploy the concept of a visual experience, nor of any mental state at all. The reflexive condition simply refers demonstratively to ways certain things have looked, or would look, where the “way” is captured by the existential portion of the content.

Second, it is far less clear that the majority of non-human animals are indeed capable of the kind of imagination being discussed. Just how large is the class of animals that can willfully call up autobiographical imagistic memories and engage in fantastical imagination, conjuring novel visual scenarios? This is a controversial question in the psychological and cognitive science literatures. Research in child-development indicates that both episodic remembering and future thinking emerge relatively late in human development, between three to five years of age (Schacter *et al*, 2007, p. 657).⁶³ Similar questions surround research on non-linguistic animals. There is some evidence that food-caching scrub jays have detailed memories of where and when food was previously cached, including what kind of food it was, suggesting a developed form of autobiographical memory (Clayton, Bussey, & Dickinson, 2003; Clayton & Dickinson, 1998). But, given their very different evolutionary histories and brain structures, it is controversial whether the cognitive process that allow them to do so are relevantly similar to human visualization.

⁶³ However, questions arise concerning children’s abilities to accurately describe their experiences at younger ages when they are perhaps already imagining things (Tulving, 2005).

In several articles and a recent book, philosopher Robert Arp (2005, 2006, 2008) makes a case for creative imagery-involving thought—what he calls ‘scenario visualization’—being a distinctively human trait. He argues that it is this capacity that enabled our human ancestors to create novel tools to solve problems in a variety of different environments. “The key feature of scenario visualization,” he suggests, is the mind’s ability to select and integrate visual images from mental modules, as well as project and transform these images in possible situations...for the purpose of solving some problem in an environment” (2006, p. 101).

While other non-human primates, and possibly some other animals, are likely capable of forming images, Arp claims that humans, have the unique ability to select among visual images, integrate and organize visual information to form coherent visual cognition, go beyond the present in order to project visual images into future scenarios, as well as transform the visual images within a variety of imagined environments (2006, p. 102).

If Arp is correct, imagination—visualization—is a quite sophisticated cognitive ability, one that enables a range of flexible, creative responses to the environment. We should not be surprised, then, if the representational contents of such states are relatively complex as well.

4. Is imagining believing?

I have argued that we can easily be wrong about what we are imagining because we lack privileged access to what we are imagining. We lack privileged access to what we are imagining because imaginative experiences have correctness conditions involving the

satisfaction of certain extra-mental conditions. We cannot, through introspection alone, determine whether those conditions are satisfied. I have tried to be explicit about what the conditions are. While an imaginer's intent plays a key role in shaping the correctness conditions of an imaginative experience, it is not sufficient to ensure that those conditions are satisfied. All of these points contribute to an understanding of how 'imagine' manages to operate as a success verb.

So far I have said nothing about the IMG operator that precedes the propositional portions of the contents described in Section 5. The IMG operator is meant to be akin to a BEL (for belief) operator. A BEL operator indicates that the proposition that follows is one that is believed, providing a notational means for distinguishing the thought from others that are not beliefs but that involve the same proposition (for instance, DES could be used to indicate that one has the attitude of desire toward the proposition that follows).

This raises the question of what work the IMG operator does in my account. Could I have not simply have used a BEL operator and argued that imagining is tantamount to forming certain complex beliefs? Interestingly, imagination as I have analyzed it is indeed closer in at least one crucial respect to belief than it is to perceptual experience. It is generally accepted that the representational content of perceptual experience can come apart from belief, in that a perceptual experience can represent things to be a way one does not believe they are (such as when one knowingly witnesses an illusion). That is, one can simultaneously (and rationally) perceptually represent that p while believing that not- p .⁶⁴

⁶⁴ A familiar example is the Muller-Lyer illusion. In this illusion, two straight lines of the same length, with arrows facing opposite directions at their ends, look to be of different lengths. This is so even if the subject knows in advance that the two lines are the same length. Having seen this illusion often enough, one may come to reliably judge, through the use of vision, that the two lines are of the same length, even while they continue to *look* to be different lengths. There is, so to speak, what one's

Indeed, this is one of the main motivations for thinking that, insofar as perception involves taking an assertoric attitude toward a proposition, it must be a different attitude than belief.

If the present account is correct, imagination does not have this feature—to imagine things just is, it seems, to believe that certain things would, or did, look a certain way, where the “way” is given by a particular portion of the content. Someone who tries to imaginatively remember the Arc de Triomphe may have the experience:

IMG MEM (there was a large white arch with thus and such visible characteristics
and the Arc de Triomphe looked roughly this way)

It does not seem that a rational person engaged in this imaginative act could simultaneously have the following belief:

BEL not-(that there was a large white arch with thus and such visible characteristics
and the Arc de Triomphe looked roughly this way)

Similarly for modal imagination, someone who tries to imagine what the Arc de Triomphe would look like (i.e., someone who hasn’t seen the Arc de Triomphe) may have an experience with the content:

IMG PM (there could be an arch with thus and such visible features AND the Arc de
Triomphe would look roughly this way)

Again it does not seem that such a person could rationally also believe that the Arc de Triomphe would not look the way her image represents it as looking.

However, I have also argued that we can willfully construct the intentional contents of our imaginative acts; if that is right, shouldn’t we be able to willfully construct an imaginative representation that one does not believe to be veridical? For instance, couldn’t

visual experience represents when taken at face value, and what one can reasonably infer based upon one’s visual experience.

someone who knew full well what the Arc de Triomphe looks like from the imaginative experience:

IMG PM (that there could be a blue sphere AND if I were to see the Arc de Triomphe it would look roughly this way).

Or

IMG MEM (that there could be a blue sphere AND I saw the Arc de Triomphe looking roughly this way).

I answer No. To willfully control the formation of imaginative experiences is not to have the ability to form any imaginative experience whatever; it is rather to choose the object of one's concern and then to represent what *it* looked or would look like in certain specified conditions (e.g., having certain properties it does not in fact have). This still allows for a large degree of the "freedom" we intuitively associate with imagination, as the specified conditions can be stipulated to differ wildly from what one believes actual. But, I claim, we are nevertheless constrained to represent it as we believe it would look in those conditions.

What do we do with the intuition (if it is one) that one can imagine the Arc de Triomphe looking like a red rubber ball, even while believing it not to look at all like that? The intuition can be recovered by holding that the content of our imaginative act in such a case is actually:

IMG PM (that there could be a red sphere AND the Arc de Triomphe, after it had been made to look like a red rubber ball, would look roughly like this).

No doubt, this is a cumbersome content to entertain; but then, it is a rather unusual and contrived sort of imaginative act.⁶⁵ What I deny is that a rational person ever has an imaginative experience that conflicts with any of his beliefs. Thus, a rational person who believes that the Arc de Triomphe does not look at all like a red rubber ball could *not* have the imaginative experience:

IMG PM (there could be a red rubber ball AND the Arc de Triomphe would look this way)

Rather, people mistake their ability to engage in the former kind of imaginative act for an ability to *willfully misrepresent*—to represent things to be a way they are believed not to be—as one would be doing were one able to form the latter experience while having the belief:

BEL not-(that there could be a red rubber ball AND if I were to see the Arc de Triomphe it would look this way)

Many of my reasons for holding this were set out in Section Three of Chapter One, where I argued against the *as present* view of imagination, and against other views of imagination that see it as essentially misrepresentational. My discussion left open the possibility that we could nevertheless, if we wished, intentionally form misrepresentational imaginative experiences. Now I am suggesting that we cannot. We can make sense of our imaginative abilities without positing an ability to form assertoric representations that conflict with our own beliefs. This allows for a simpler account of what sensory imagination really is—a representation of how things looked, or would look—and avoids the need to posit a

⁶⁵ It is also questionable whether this is a veridical case of modal imagination, since it is doubtful that the Arc de Triomphe could be transformed into a rubber ball. See Chapter Three for an extended discussion of cases where IMG MOD experiences involve entertaining metaphysically impossible propositions.

distinctive “imaginative” attitude, or a separate cognitive “box” (akin to a “belief box”) where propositions involved in imagination are kept quarantined from those that are believed (as against the claims of “single code” theorists such as Currie and Ravenscroft (2002), and Nichols and Stich, (2000)).⁶⁶ Does this mean that imagining can take place entirely within the “belief box”? Seeing isn’t believing, but could imagining be?

My answer is a qualified no. As far as accounting for the representational content and correctness conditions of imaginative experience goes, I think these could equally well have been captured with the use of a BEL operator. I have used the IMG, however, because it seems one could believe the propositions that follow it, in the examples above, without actually *imagining* anything—mainly because it seems one could form those beliefs without forming an *image*. Here I have in mind the amnesiac patients studied by Rubin and Greenberg, who lack the ability to form visual images. Such patients likely have beliefs about their past experiences, gained through subsequent conversations with friends and family (so long as they do not have severe retrograde amnesia as well, as most Rubin and Greenberg studied did not). Yet they are unable to form the kind of imagistic memories that we naturally associate with genuine autobiographical episodic memory. Thus, despite the fact that we cannot simultaneously rationally have an imaginative experience with a representational content p and a complex belief with the content not- p , it *does* seem that one could believe that p , while not having the ability to imagine that p , and this suggests that imagining is not strictly equivalent to believing.

To be clear, I still hold that the representational (or intentional) content of an act of imagination can be exactly the same as that of a (suitably complex) belief. What differences

⁶⁶ See Chapter Four for a detailed examination of Nichols and Stich’s (2000) and Currie and Ravenscroft’s (2002) theories in the context of propositional imagination.

exist between the imaginative experience and belief above will not, if I am right, be differences in representational content, nor even in the final role such states play in guiding practical reasoning. Rather, they will be differences in, perhaps, the (non-intentional) phenomenal character of the experiences, in the kinds of abilities deployed in generating such contents, in the way the information contained in the representation is structured and accessed, or even in the format of the representations. In other words, there may be *more* to imagination than what I have described; but that “more” will not include other representational features.

The IMG operator can also play an important role in indicating that distinctive cognitive mechanisms are involved in imagination that are not involved in mere belief. For there is overwhelming empirical evidence⁶⁷ that neural populations selectively involved in perception are similarly activated in imagery (see Currie and Ravenscroft (2002, Ch. 4) and Goldman (2006b) for detailed discussions of this work). Including an IMG operator indicates that a distinctive cognitive ability, linked neurally and developmentally to one’s perceptual abilities, is involved in the generation of mental images. Indeed, perceptual deficits are often mirrored in imagery. Sirigu *et al.* (1995) found that deficits in a patient’s ability to move her fingers were attended by similar deficits in her ability to imagine moving her fingers. Corresponding data has been found regarding visual imagery and vision (Brooks, 1967): patients who, due to brain lesions, neglect a certain region of the visual field will

⁶⁷ Jeannerod and Pacherie (2004, p. 130)(2004, p. 130) discuss results from neuro-imaging studies showing activity in a common neural network activated during both (mere) motor imagery and action preparation. Currie and Ravenscroft (2002) cite a wealth of similar data. More recently, Slotnick *et al.* (2005) found similar visual imagery and perception tasks to evoke cortical activity with the same precise visual field topography (‘retinotopy’), and to cause common effects in motion processing region MT+. Kosslyn and Thomson (2003) present a meta-analysis of a variety of neuroimaging studies with similar findings.

neglect that same region in visual imagination.⁶⁸ It is therefore plausible that one could lose the ability to have imaginative experiences with certain contents while maintaining the ability to have beliefs with those contents, as would seem to be the case with the patients Rubin and Greenberg discuss.

I end this chapter with a last speculation, which is that imagination is distinctive in being constituted by the conjunction of representational contents generated through the interplay of two different cognitive systems: perceptual memory centers that generate the “existential” aspect of the content, and “central processing” (Fodor, 1983) that places an additional condition upon the existential component. In Chapter Five, I look closely at the evidence for there being two such systems at work in imagination, connecting their interplay to the phenomenological sense of thought-agency. Next, however, I will apply the theory of sensory imagination developed so far to a specific issue in contemporary philosophy; this is the question of whether, and to what degree, imagination serves as a guide to possibility.

⁶⁸ Similar studies are discussed in more detail in Chapter Five.

CHAPTER THREE:

The Appearance of Possibilities

1. Introduction

Here is a familiar idea: imagination stands to the possible as perception stands to the actual. Just as perception represents how the world is, imagination represents how the world may be, or could have been. The basic claim is clear enough—our warrant for accepting it (if we do) much less so. In this chapter I try to shed some light on the hoary doctrine that imagination is a “guide” to possibility, by applying the theory of sensory imagination developed in chapters one and two to a contemporary debate in modal epistemology. My hope is that the theory itself, along with some of its motivations, will be further elucidated in the process.

The debate in question grows out of Kripke’s (1980) famous arguments for *a posteriori* necessities and *a priori* contingencies. One of the most remarkable features of these arguments is their ability to flip (previously) common modal intuitions on their heads, by eliciting further intuitions we didn’t know we had. This leads to a familiar tension in his work. On the one hand, Kripke seems to establish through more or less intuitive means that certain things traditionally thought possible—and probably still thought possible by most non-philosophers—are not possible after all. On the other, this conclusion threatens to undermine the very intuitions of possibility and necessity that seem to motivate it in the first place. For if our intuitions about possibility can be as badly mistaken Kripke showed them to be, why think our new ones are reliable?

The skepticism is perhaps unwarranted if Kripke has provided a plausible modal epistemology, one which shows us how modal appearances can be misleading in some cases, and reliable in others. Whether he has, and to what degree he has, remain matters of considerable controversy. Usually these questions are thought to turn on the issue of what links, if any, hold between imaginability (or conceivability) and possibility. Some find it obvious that Kripke's work impugns imagination and conception-based evidence for modal claims; others insist that he ultimately vindicates the power of imagination to provide modal insight, once certain common errors are corrected for.

Alex Byrne (2007) appears to fall somewhere between these two groups. According to Byrne, when Kripke says that something is imaginable, he is not endorsing imaginability as a guide to possibility, but rather is "simply making various claims about what could and couldn't have been the case, which strike him as intuitively correct." That is, Kripke is not suggesting that he has determined the thing is possible *through* imagining it. His intuitions simply are what they are, Byrne suggests, and when Kripke says something is imaginable, or that it *cannot* be imagined, he is just voicing the intuition that it is possible, or not possible, without comment on its provenance. Thus, to question the coherence of Kripke's modal epistemology—and, in particular, his appeal to imaginability as a guide to possibility—is to miss the point, since Kripke offers no modal epistemology at all.

I agree with Byrne that Kripke does not offer imaginability as a *guide* to possibility. When Kripke says something is imaginable, it is directly implied that the thing is possible, just as it is implied that *p* is true when someone is said to know that *p*, or that a fox is there when Sam is said to see a fox. For Kripke, imaginability is no more a *guide* to possibility than knowledge is a *guide* to truth. This helps explain a number of his comments that I will

come to shortly. Yet, I do not agree that Kripke offers no modal epistemology at all. I argue here that he (only implicitly) offers *apparent imaginability* as a defeasible “guide” to possibility, in roughly the way that perceptual experience is a defeasible guide—if not “evidence”⁶⁹—to what is around us.

What could the difference between apparent imaginability and imaginability *tout court* amount to? Well, describing the difference obviously requires having in place a clear distinction between imagining *X* and merely seeming, or appearing to imagine *X*. In chapters one and two I gave an explicit account of what I think this distinction comes to; in this chapter I say how I think it applies to Kripke’s arguments. I’ll proceed by linking the notion of modal appearances at play in Kripke’s work to his appeal to qualitative epistemic situations, and linking qualitative epistemic situations to the conception of modal imagination I outlined in Chapter Two.

I argue that this assimilation makes sense of a range of puzzles surrounding Kripke’s modal epistemology, and provides a more plausible extrapolation and extension of this epistemology than some other well-known contenders. Finally, I draw out the implications this assimilation has for Kripke’s arguments against mind/brain identity theories.

2. Factive Imagination?

While it is generally accepted that Kripke works with an appearance/reality distinction when considering counterfactual situations—things can seem possible that are not possible—what these “seemings” amount to, cognitively, for those considering the relevant

⁶⁹ Intuitively, perceptual experience provides too direct and reliable a source of information about the immediate environment for it to be considered mere “evidence” for what is around us. Perceptual experience is nevertheless a “guide” to what is around us insofar as it warrants our claims about what is around us.

possibilities, has been poorly understood. We can see this by looking closely at the way Kripke uses the word ‘imagine’. For Kripke, everything imaginable (and not merely apparently imaginable) is possible. For, as Byrne points out, if Kripke were offering imaginability as mere evidence for—but not as entailing—possibility, then his flat-footed claims that we *cannot* imagine certain things seem rather bizarre. For instance, Kripke says that, if the molecular theory of matter is correct, then for any particular object “the one thing we *cannot imagine* happening to this thing is that *it*, given that it is composed of molecules, should still have existed and not have been composed of molecules” (1980, p. 126-7, first emphasis mine).

Similarly, he suggests, while it may *seem* that we can imagine a certain wooden table being made of ice, “this is *not* to imagine *this* table as being made of wood or ice, but rather it is to imagine another table, *resembling* this one in all external details, made of...ice” (1980, p. 114). Byrne’s response to these claims seems entirely legitimate: “Offhand,” says Byrne:

It seems pretty easy to imagine this [wooden] table as made of ice. Is Kripke claiming that I have *misidentified* the content of my imagining...That is a quite remarkable charge, requiring some supporting argument, and Kripke provides none; indeed, he shows no sign of recognizing that he is saying anything outlandish about the limits of imagination (2007, p. 126).

If imagining is simply a matter of entertaining a proposition that attributes a familiar property to a familiar object—being made of ice to a table, or being inanimate to Nixon—then Kripke’s claims that we *cannot* imagine thus and such seem rather peculiar. Why would we be incapable of entertaining these propositions? After all, Kripke never suggests that we

cannot *believe* impossible propositions. So then, why couldn't we just as easily imagine them? (Note: This question presumes, in line with the standard view of things, that propositions are the objects of imagination. This is not my view of the matter, however—nor, I argue here, is it Kripke's.)

A common response to this problem, also considered by Byrne, is to hold that in cases where we take ourselves to be imagining the things Kripke says we cannot imagine, we are confused about the proposition that we are in fact entertaining. Stoljar (2006, pp. 74-76) calls this “proposition confusion.” On this fairly standard view, it is not that we *cannot* imaginatively entertain certain propositions, it is rather that the intuitive sense we perhaps had that such propositions expressed genuine possibilities is grounded in the fact that we were actually imagining some other proposition that really *is* possible. Perhaps this is what is going on when Kripke says that, when it seems one can imagine a certain wooden table having been made of ice, one is really only imagining an ice table resembling that table “in external details”. The advocate of “proposition confusion” formalizes this claim in the following way: the person who thought he was imagining that the wood table could have been made of ice was actually entertaining a different, but related proposition, namely, that there could be an ice table superficially similar to the wood table. Once we recognize that the latter proposition represents a genuine possibility and the former does not, we will be inclined to say it was the latter we were initially imagining. The question of whether or not we can imagine impossible propositions is side-stepped, since the claim is that one mistook one's imagining of a possible proposition for an imagining of a certain other (impossible) one.

This response faces at least two major worries. The first is that it requires us to be implausibly inept at determining the propositions we are imaginatively entertaining. Byrne makes this point, supposing himself to be standing in front of the very wood table at the center of controversy:

Mistakes about the content of one's imaginings can happen, no doubt, but why would they happen when I am actually eyeballing the table itself, and presumably in the best possible position to imagine something about *it*? As Wittgenstein pointed out (1969, p. 39), we don't normally think that the attempt to imagine King's College on fire is a spotty, hit-or-miss affair, with the imagination often only connecting with the proposition that some similar building is on fire (2007, p. 126).

The advocate of the proposition confusion interpretation of Kripke needs to say why we should be so confused about the propositions we are entertaining. Certainly, if propositional attitude psychology is to explain human behavior, it will not do to simply hold that we are very often unreliable judges of the propositions we are entertaining.

The second problem is that the proposition confusion strategy clashes with Kripke's claims that we cannot imagine certain things. When Kripke says that we cannot imagine a particular molecularly-constituted table not having been made of molecules, or the Queen being born from parents other than her actual parents, he clearly suggests that, even after whatever relevant confusions have been cleared up, we *still* cannot imagine such a thing. But then, why we would not be able to imagine such things *if* imagining consists simply in entertaining certain propositions? It is a mark against the proposition confusion approach that it has no clear way of explaining this current in Kripke.

With these worries in mind, we can look more closely at Byrne's analysis of the situation. He points out that Kripke uses 'possible world' more or less interchangeably with 'counterfactual situation,' or simply 'situation'; "talk of situations," Byrne suggests, "is talk of what *could have been* the case." So, Byrne says:

When Kripke says that we can't imagine a situation in which the Queen is the biological daughter of Mr. and Mrs. Truman, 'imagine' could easily be otiose: we can't imagine (or "think of") such a situation, not because of limits to our imagination, but because there is no such situation to imagine (2007, p. 127).

Byrne rejects the idea that, in Kripke, imagining amounts to entertaining propositions; it is not a limit on some such faculty that prevents us from imagining the situation where the Queen is the Trumans' daughter. Rather, it is a limitation on the counterfactual situations that actually exist that stands in the way—just as it is a limit on what *actually* exists, not a limitation on the faculty of perception, that prevents us from seeing Santa Clause.

While the last claim extrapolates a bit from what Byrne actually says, it fits with another way Byrne makes his point. Byrne says that we should interpret Kripke's claim that we can't "think of" a counterfactual situation where Nixon is inanimate as being akin to the claim that one "can't 'think of' a highest prime" (2007, p. 128). We cannot think of a highest prime because there *is* no highest prime (otherwise one obviously *could* think of it, if perhaps only accidentally). Moreover, one cannot think of the highest prime even though one can easily enough *believe* oneself to be thinking of the highest prime.

Again there is a kind of factivity suggested by Kripke's "think of" and "imagine" locutions. If one can think of the counterfactual situation, or imagine it, then there is such a counterfactual situation. Thus, the ability to think of or imagine such situations depends not merely on our ability to entertain some relevant proposition—it depends also upon whether that proposition represents a genuine counterfactual situation. If that is right, then it offers a clean explanation of Kripke's claim that we cannot imagine thus and such, an explanation not implying that we cannot entertain certain seemingly basic propositions, or tell what

proposition we are really entertaining when we entertain it (even if it leaves completely open the question of how the propositions we entertain manage to “connect” with the counterfactual situations they target).

Of course, this move towards factivity raises some questions. First, there is the thorny issue of what it means for a (merely) counterfactual situation to be “genuine” or to “actually exist.” This I propose to set aside; our concern here is not with the metaphysics of modality, but with modal psychology and epistemology.⁷⁰ More to our purposes, imagining a situation on Byrne’s view of Kripke apparently requires more than imaginatively entertaining a certain proposition. It requires that there *be* such a situation to imagine as well. But Byrne does not say whether this means propositions, or something else, are the objects of imagination. Nor does he indicate how the purely mental side of imagining represents, or otherwise interacts with, the various counterfactual situations that do and do not exist.

Maybe this is because he thinks imaginability in Kripke is “otiose”, simply irrelevant to the issue of how or why something is possible. But I think that, if we push further here, we will see that Kripke *does* have a kind of modal epistemology in play after all, contrary to Byrne’s central claim.

So consider again Byrne’s claim that we cannot imagine certain situations “not because of limits to our imagination, but because there is no such situation to imagine.” An analogous claim about vision would appear to be: we can’t see the Tooth Fairy not because of limits to our visual system, but because there is no Tooth Fairy to be seen. Nevertheless,

⁷⁰ Yes, one’s epistemology will likely be affected by one’s metaphysics, and vice versa. I simply propose to approach the present issues from the angle of epistemology, letting the metaphysical chips fall where they may, for now.

we could *seem* to see the Tooth Fairy—and to suggest that we could not *would* be to put a restriction on our visual system. Moreover, seeming to see the Tooth Fairy on one's window sill is warrant for believing that Tooth Fairy is on one's window sill (especially if the relevant experts (one's parents) have told you that there exists such a being).

Can these analogies be carried over to the case of imagination? Can we be in mental states—"imaginative experiences"—in virtue of which we *seem* to be imagining a certain counterfactual situation? And can being in such states provide warrant for believing that there is such a counterfactual situation—even if, sometimes, there is not? These questions motivate the section to come.

3. Illusions of Possibility

Even if we can only imagine what is metaphysically possible, in the same way that we can only see what exists, there is room for a mental state in which perhaps impossible things seem possible, analogous to the illusory and hallucinatory perceptual states we enter into when various nonexistent things or properties seem present before us. I've called such imaginative states "imaginative experiences". Is there any evidence that *Kripke* accepts this analogy, whereby imaginative experiences provide (defeasible) *appearances* of possibility? We can begin by considering the following passage, in which Kripke discusses imagining substances whose atomic number is not 79:

Any world in which we imagine a substance which does not have these properties [i.e., atomic number 79] is a world in which we imagine a substance which is not gold, provided that these properties form the basis of what the substance is (1980, p. 125).

Note that Kripke does not deny that people can *believe* themselves to be imagining gold—or *seem* to be imagining gold—that lacks atomic number 79.⁷¹ On the contrary, he is merely pointing out that, *regardless of what one believes or seems to be imagining*, if one is imagining a substance whose atomic number is not 79, then one is not imagining gold.

He makes a similar claim about a case where something other than electrical discharge creates a flash in the night sky: “Here too, I am inclined to say, when we imagine this, we imagine something with all the visual appearances of lightning, but which is not, in fact, lightning” (1980, p. 132). Again, he does not deny that people have taken themselves to be imagining lightning that is not electrical discharge—he simply claims that when we imagine flashes in the night sky that are not electrical discharge, we are not imagining lightning—even if we declare to ourselves that we are.

Note how clean the analogy to vision is: if one were to *see* flashes in the night sky that were not electrical discharge, one would not be seeing lightning, even if one thought one was. Similarly, Kripke seems to suggest, if one *imagines* flashes in the night sky that are not electrical discharge, then one is not imagining lightning, even if one thinks one is.

So, there is a fact of the matter, according to Kripke, concerning what one is imagining that, apparently, is not fixed by one’s beliefs about what one is imagining—just as there is a fact of the matter concerning what one is seeing that is not fixed by one’s beliefs about what one is seeing. Supposing we accept this analogy, there is room for a notion of *appearances* in both cases: were one to unwittingly see “fool’s lightning” illuminating the night sky, one would be enjoying a perceptual experience that (non-veridically) represents

⁷¹ Though some have been inspired by Kripkean ideas to suppose that one *cannot* believe impossible propositions—and that believability therefore entails possibility. Obviously, this view has its own baggage to carry.

lightning.⁷² That is, one could have a visual experience *as of* lightning while in fact *seeing* only fool's lightning. Similarly, when one tries to imagine lightning that is not electrical discharge, one may have an imaginative experience *as of* lightning that is not electrical discharge, while in fact only imagining fool's lightning—or, perhaps, nothing at all.

In the case of vision, there is a relatively uncontroversial way, appealing to causal relations, of specifying what, if anything, I see when I don't see what I believe I am seeing. In the case of imagination, we might accept that we were not imagining what we thought we were, while still wondering how to determine what (if anything) it was that was actually imagined. There are at least two ways to approach this issue. First, one could hold that all non-veridical imagination is akin to hallucination, being *of* no real thing at all. Alternatively, one could hold that non-veridical imagination is akin to illusory experience, and then suggest a diagnostic for determining what was in fact imagined from case to case. This might involve asking the subject to fill in details about her imaginative project to see if they suggest a particular alternative target.

It is worth noting that a similar difficulty arises for vision as well: there will inevitably be a gray area between illusion and hallucination where it is not clear whether we should say a subject is having a radically illusory experience of an existent thing *X* or, instead, merely hallucinating a non-existent *Y*.

Note that if Kripke did *not* accept a seems/is distinction in the case of imagination (and did not hold that we can only imagine what is possible), he could have just said: we can imagine gold whose atomic number is not 79, and lightning that is not electrical discharge—but, alas, such things are not possible. But not only is a seems/is distinction suggested by

⁷² Or, if one thinks the property of being lightning is not represented in perception, one would have a false belief about what one is seeing.

Kripke's apparent "success" verb use of 'imagine', he explicitly appeals to modal appearances in several places—both when he wants to explain away certain appearances that he thinks are misleading, and when arguing that appearances weigh against mind/brain identity theories. In both instances he appeals to the notion of a 'qualitatively identical epistemic situation' ("QIES") in order to show how two different scenarios—one possible, the other impossible—can have the same appearance of possibility. Or so I will now argue.

Kripke first introduces the notion of a QIES to explain away the illusory possibility that a certain wood table could have been made of ice:

What, then, does the intuition that the table might have turned out to have been made of ice or of anything else...amount to? I think that it means simply that there might have been *a table* looking and feeling just like this one...which was in fact made of ice. In other words, I (or some conscious being) could have been *qualitatively in the same epistemic situation* that in fact obtains, I could have the same sensory evidence that I in fact have, about *a table* which was made of ice (1980, p.142, emphasis in original).

For Kripke, two "situations" are qualitatively identical if (roughly) they involve objects with the same external appearances.⁷³ It seemed that a certain wood table could have been made of ice because there really *could* be an ice table "looking and feeling just like" the wood table in question. That there could be such a table is meant to explain away the sense that something impossible (the wood table having been made of ice) is in fact possible. And,

⁷³ It is not always clear what Kripke means by "appearances"—are they superficial qualities possessed by objects, or mental states in virtue of which certain qualities appear to be instantiated in the environment? Kripke sometimes glosses being in QIESs as having "the same sensory evidence" (p. 142) in two situations, which suggests the latter interpretation. This is evident in his discussion of heat, where he describes the relevant QIES as one where someone "sensed a phenomenon in the same way we sense heat, that is, feels it by means of its production of the sensation we can 'the sensation of heat'" (p. 151). Here, heat's "appearance" just is a certain sensation—which is obviously not a property of heat itself. Yet in other places he speaks simply of two objects sharing the same "surface properties".

indeed, this style of explanation *does* alleviate, for many, the sense that there was a genuine possibility in the vicinity. As Stephen Yablo (2006) notes:

Philosophers have been telling us for centuries that this or that common impression is false. And we have for centuries been shrugging them off. What makes Kripke special is that he gets you to *agree* that you are making the mistake he describes (2006, p. XX).

Yablo cites the ice table example as paradigm case of Kripke offering not merely an explanation of why we *might* be wrong about a certain possibility, but, more amazingly, giving an explanation that one can't help but believe (thereby meeting Yablo's "Psychoanalytic Standard"⁷⁴).

Doubtless Kripke's explanation seems plausible. But *why*, we might pause to ask, should the fact that there could have been an ice table that looks just like the wooden one alleviate the sense that the wooden one could have been made of ice? Why is the possibility of such a table relevant at all? It's not as though we formerly thought that ice tables with certain appearances were *impossible*...so how is affirming their possibility relevant to our putatively mistaken intuitions about *this* wood table?

Similar questions can be asked for Kripke's other examples involving "qualitatively identical evidential situations," where, he says, "something like counterpart theory" is applicable in revealing what really is possible (1980, p. 142).⁷⁵ The false statement that

⁷⁴ Yablo's Psychoanalytic Standard, which he thinks is met by Kripke's explanations of mistaken modal intuitions, is: "Assuming the conceiver is not too self-deceived or resistant, POSS F explains E's seeming possibility only if he/she does or would accept it as an explanation, and accept that his/her intuition testifies at best to F's possibility, not E's" (2006, p. XX).

⁷⁵ Kripke famously lampoons the epistemology of counterpart theory as involving peering through "telescopes" at possible worlds in order to identify, by their superficial characteristics, relevant counterpart objects (to "see", for example, whether or not there is a person in another possible world who closely enough resembles Nixon, yet who was never president, as a means of determining whether Nixon could have not won the presidency). It is interesting to note that the very method he mocks as a procedure for determining possibilities is what he appeals to in order to explain away mistaken intuitions; the suggestion is that we *can* determine purely

Hesperus might not have been Phosphorus is replaced by the statement: “two distinct bodies might have occupied, in the morning and the evening, respectively, the very positions actually occupied by Hesperus-Phosphorus-Venus” (p. 143). And “the loose and inaccurate statement that gold might have turned out to be a compound,” is replaced by: “it is logically possible that there should have been a compound with all the properties originally known to hold of gold” (*ibid.*). Again, we may accept these explanations, yet still ask: why are *these* the relevant alternative possibilities?

To answer this question, a more explicit connection between QIES and our *means* of determining whether something is possible must be drawn. As a first step, consider cases of perceptual confusion. The situations where we would most likely perceptually mistake an ice table for a wood table (or fool’s gold for gold, fool’s lightning for lightning, etc.) are situations where the two things share the same outward appearances. In the case of perception, it is obvious why an object’s having the same outward appearance as another is relevant to explaining why we confuse them—outward appearances are all we have access to with perception.

The question before us is why shared outward appearances should matter in the case of mistaken possibilities. Here is a guess: we often use sensory imagination (i.e., imaginative experience) as a guide to possibility, and sensory imagination always, at a minimum, represents the superficial, outwardly apparent properties of various possible objects. When attempts to imagine different counterfactual situations involve representing

qualitative possibilities by peering through mental telescopes, we simply must resist making modal conclusions about *particular* objects, natural kinds, and things picked out by rigid designators, when looking through such “telescopes”.

the same perceptible properties—involve the same visual image—we risk mistaking one such situation for another.

If this is right, then the reason QIES provide intuitive relief may be the following: when we are not imagining what we thought we were imagining, there will usually be another imaginative project incorporating the same mental image in which a genuine possibility is represented. And this may be enough to explain the sense that one was veridically representing something in one's original, unsuccessful, imaginative project.

Note that this is not to say that, in the case where it seems a wood table could have been made of ice, one was having an imaginative experience as of *some other* ice table all along. That would be to fall into the problems of the 'proposition confusion' strategy above, where one lacks reliable epistemic access to the propositions one is entertaining. Rather, it is to suggest that one's imaginative experience as of this very wood table being made of ice was non-veridical, yet *phenomenally similar to* another possible imaginative experience: the experience of imagining a superficially similar table that is made of ice. So, one was not wrong about the propositions being entertained—one really had an imaginative experience as of this very wood table being made of ice. It is just that the experience did not represent a genuine counterfactual situation. Its seeming to represent a genuine possibility can be explained by the fact that there are other imaginative experiences that involve the same or similar appearance of possibility (due to their incorporating the same image) that do represent genuine possibilities.

This is to reverse the normal order of explanation. Normally it is supposed that one act of imagining (i.e., a *mental* action of some kind) is confused with another because the non-mental things at which these acts are aimed share certain superficial properties. I

suggest that it works the other way around: one possibility is mistaken for another (im)possibility because the mental—imaginative—states we use to imagine both possibilities share representational properties. That is, they share the same image.

So, my suggestion is ultimately quite simple: perceptual confusions result from sameness or similarity in perceptual experience, while (at least some) modal confusions result from sameness or similarity in imaginative experiences—where the relevant similarities in the case of imagination are grounded in the use of the same image-type. Again, this helps explain why Kripke is so often concerned with the superficial, perceptible properties of possible objects when explaining away mistaken intuitions. Moreover, this approach respects the common intuition—supported by the empirical work discussed in Chapter One—that sensory imagination is at work in the day-to-day involvement of modal determinations involved in anticipation and planning. For it is natural to think that—rightly or wrongly—we might then also draw on sensory imagination when considering the more unusual counterfactual situations relevant to Kripke’s arguments.

4. Are QIES Necessary?

Before connecting QIES more explicitly to my account of Modal Imagination, a brief detour is in order. Insofar as I am linking Kripkean modal reasoning to sensory imagination via his notion of QIES, it is worth asking whether QIES really play the central role in his modal epistemology that I am suggesting. As we’ve seen, there are many places where Kripke specifically appeals to QIES, but he also mentions another, seemingly related, strategy for explaining away mistaken modal intuitions. This strategy is called “Textbook

Kripkeanism” by Yablo (1999)⁷⁶ and is critically discussed by Bealer (1994). Textbook Kripkeanism appeals to our putative weakness for unwittingly substituting *reference-fixing descriptions* for the rigid designators flanking an identity sign, thereby revealing the contingent proposition responsible for the mistaken modal intuition. For instance, by rephrasing ‘Hesperus is Phosphorus’ as, ‘the first star seen in morning is the first star seen in the evening,’ we reveal the latter as the truly contingent statement that accounts for the common intuition of contingency concerning the former.

A limitation of this strategy is that, as Kripke himself recognizes (1980, p. 143), it only provides a clear method for explaining away mistaken intuitions regarding *identity statements*. When it comes to cases such as the ice table example, or the scientific bases of natural kinds, we must revert to what Kripke calls “our general paradigm,” which “is to redescribe both the prior evidence and the statement qualitatively and claim they are only contingently related” (1980, p. 143). It is a special virtue of the appeal to QIES that it can explain away mistaken intuitions regarding both identity statements *and* constitution claims (e.g. the ice table example), the latter of which leave Textbook Kripkeanism at a loss. That is, the apparent contingency of ‘Hesperus is Phosphorus’ can be explained by the fact that someone looking at Hesperus could be in a QIES when looking at a star that is not Phosphorus. Yet, it is unclear how the mistaken intuition that a certain wood table could

⁷⁶ Yablo (1999) describes Textbook Kripkeanism as a strategy that explains away various mistaken intuitions concerning a proposition E by appealing to a proposition E* that “maps out the way the proposition that E is presented in thought; it is, for short, a presentation of E. The usual sort of presentation takes name-like expressions in E and replaces them with descriptive and/or demonstrative phrases that, as Kripke says, fixes their reference” (p. 456).

have been made of ice (or those concerning natural kinds) might be explained away by substituting reference-fixing descriptions.⁷⁷

Finally, one might respond that the *most* general way of understanding Kripke's strategy for explaining away mistaken intuitions would simply be to hold that whenever a necessary statement is thought to be contingent, there is always *some other* contingent statement lurking that, once recognized, we will allow is was what we had in mind all along—without commitment to the nature of that contingent statement from case to case.⁷⁸ However, this kind of strategy provides no epistemological comfort to anyone who ultimately wants to make substantive claims of contingency (e.g. Kripke, esp. with regard to mind/brain identities), since it is always open to the opponent of such contingencies to hold that it is some other statement that one has in mind. Moreover, it provides no means of predicting what sorts of errors we should expect in modal reasoning—it only holds that errors may abound—which calls into question whether the supposed metaphysical modality is something we have any reliable insight into whatsoever.

In focusing on Kripke's appeal to QIES, I am not denying that there are other forces at work in Kripke's (and our own) modal reasoning, or suggesting that QIES can explain away all putatively mistaken modal intuitions (they can't⁷⁹). I am denying, however, that there is a perfect, entirely general strategy available to replace QIES, and that, for this reason,

⁷⁷ For other reasons why Textbook Kripkeanism cannot be the principal strategy for explaining away mistaken intuitions, see Bealer (1994, esp. 194-6) and Yablo (2006).

⁷⁸ This is close to following thesis "O" that Yablo concludes might justify conceivability evidence as a guide to possibility: "Carefully handled, conceivability evidence can be trusted, for when impossible E seems possible, that will generally be because of distorting factors that we can discover and control for" (2006, p. XX).

⁷⁹ It seems doubtful that QIES can account for mistaken mathematical intuitions—e.g., that a certain result could have come out differently. See also Yablo's discussion of heat (2006, p. X-X).

they play a central, even necessary, role in Kripke's model epistemology.⁸⁰ This warrants drawing as clear a picture of this notion—and its relation to sensory imagination—as possible.

5. *From QIES to IMG PM*

I now turn to the task of explicitly linking Kripke's appeal to QIES to the account of Modal Imagination developed in Chapter Two. Recall that any act of modally imagining some object or situation will involve a complex content, consisting in the conjunction of two modal components. The first conjunct accounts for the contribution of the "image," while the second further specifies what the image represents. I'll stick with Kripke's ice table example in showing what the contents of the relevant imaginative experiences would be. When a person tries to imagine a certain (demonstrated) wood table with superficial features q,r,s being made of ice, the experience is:

IMG PM (there could be a table with q,r,s AND this table made of ice would look roughly like this).

The last 'this' of the second condition again refers to the way the first "imagistic" portion of the content represents a possible ice table as looking. Now, if Kripke's claims about the essences of tables are correct—in particular, if a certain wood table could not possibly have been made from anything other than the piece of wood it was in fact made from—then the second condition in the above content is not met. It is not true that this table

⁸⁰ Yablo (2006) agrees with the centrality of QIES as a means for explaining away mistaken modal intuitions. He calls the appeal to QIES the "epistemic counterpart model" (EQ). According to EQ, when there is an illusion that A could be Q, "it seems possible for A to be Q because it really is possible for A* to be Q, where A* is a facsimile of A" (X is a facsimile of Y iff X would be "superficially indistinguishable" from Y *for us* (and not merely some other possible non-human perceiver)). A main argument of Yablo's (2006) is that Kripke's account of some mistaken intuitions (e.g. that heat could have not been *mke*) fails because they cannot be explained away in terms of EQ.

when made of ice would look roughly the way my image represents a table as looking, because there is no counterfactual situation where this very table is made of ice.⁸¹ Hence there is no counterfactual situation where it—the ice/wooden table—looks any way at all.

What is possible—and is imaginable—is *an* ice table with the superficial features *q,r,s*. And one can imagine such a table with the following experience, which, on my account, amounts to imagining an epistemic situation qualitatively identical to the one in which one is looking at a certain wood table:

IMG PM (there could be a table⁸² with *q,r,s* AND an ice table would look roughly this way)

This is the imaginative experience whereby one imagines *an* ice table with the superficial characteristics of the wooden table described above. It features the same image—the same existential content—as the non-veridical imaginative experience above, yet is a veridical case of imagination, indeed of “veridical imagining,” since in this case both conditions are (presumably) met. As discussed in Chapter Two, when both conditions are met, it is the object referred to by the second conjunct of the content that is imagined.

The element common to the two acts of imagining is their existential, “imagistic” component. This suggests a further conclusion: any two counterfactual situations one tries to imagine with the same type of image will be imagined QIESs. To test this claim, consider a situation where what looks to be a wood table with *q,r,s* is actually a *trompe-l’oeil* painting on a flat canvas. The qualities present *in that situation itself* will be very different than the qualities present in a situation where a normal wood table is in fact present (for instance, the

⁸¹ See Section 6 below for important elaboration on this point.

⁸² For ease of exposition, I assume here that the property of table-hood could be represented by the imagistic portion of one’s experience (and, by extension, that it could be represented by perceptual experience). A geometric description could replace “table” if one thinks table-hood is not representable in perceptual/imagistic experience.

actual shape of the object seen is very different). Nevertheless, one might use the same mental image to imagine both scenarios. Supposing it seems possible for the wood table to have been a *trompe-l'oeil* painting, a plausible way of explaining this intuition away is a Kripkean appeal to a QIES: one could have been in a QIES where there was a *trompe-l'oeil* painting that looked to have q,r,s . Notice that the reason this appeal works is *not* because actual external qualities of objects are duplicated in the two situation (they're not), but because we use the same image to imagine the *trompe-l'oeil* situation that we would to imagine a situation involving the wood table itself.

So, the notion of a QIES appears *secondary* to, insofar as it is generated out of, sensory imaginings that involve the same mental image. That is, epistemic situations are “qualitatively identical” in Kripke’s sense not because they necessarily involve objects with the same superficial properties, but because we use the same type of image—an image that represents the same superficial properties—to imagine them. When we combine this reduction of QIES to the use of the same (type of) image to imagine different situations with our use (rightly or wrongly) of sensory imagination when considering possibilities, we begin to see more clearly why the appeal to QIES is so explanatorily powerful.

6. A digression on two technical issues

Before moving on, however, there are two technical worries that must be addressed concerning the “non-veridical” imaginative experience described above. Here it is again: IMG PM (there could be a table with q,r,s AND this table made of ice would look roughly like this).

The second conjunct of the content may seem equivalent to a conditional of the form: ‘if this (wood) table were made of ice, it would look roughly this way.’ If we accept that the second conjunct is indeed equivalent to this conditional, some potential problems emerge. First, note that the antecedent of this conditional represents a metaphysical impossibility—namely, that a certain ice table is (counterfactually) made of wood. Some have argued that there are no metaphysically impossible propositions (since they identify propositions with possible worlds); others argue that it is incoherent to believe a metaphysical impossibility, and that believability requires possibility. I ascribe to neither view. Following Roy Sorenson (1996), I think it is perfectly coherent to believe a metaphysical impossibility—to believe, for instance, that gold has atomic number 78.⁸³ Why else would Kripke’s arguments that gold could not possibly have had an atomic number other than 79, etc., have been met with such surprise? Did they really not conflict with anyone’s prior beliefs?

The second, more pressing, technical issue involves the threat of the conditional seemingly implicit in the second half of the content being vacuously true. According to standard truth-functional logic, all conditionals with false premises are true (if “trivially” or “vacuously”). As already remarked, is plausible to think that the statement:

[1]: ‘this (wooden) table made of ice would look roughly this way’

⁸³ Sorenson has an interesting “transcendental” argument to the effect that metaphysical impossibilities are believable. Roughly: if we cannot believe metaphysical impossibilities, then we necessarily cannot. On the view that we cannot believe necessary falsehoods, this would entail that it is not possible to believe that we can believe metaphysically impossible propositions. If so, the defender of the view that we *can* believe impossible propositions cannot be accused of having a false belief. “If the impossibilist [who denies the believability of impossibilities] can get into the belief relation with ‘No one can believe an impossibility’,” writes Sorenson, “then I can get into the belief relation with its negation,” since “the negation of any meaningful sentence must itself be meaningful” (1996, p. 247).

As Sorenson also notes, it is one thing for the impossibilist to say that a person who claims he believes that Hesperus is not Phosphorus is “confused” about what he believes (as there is a psychologically pleasing alternative account of what is believed is available), quite another to say that the person who claims he believes that metaphysical impossibilities are believable is similarly confused—what simple strategy is there to make it plausible that one does not even *believe* this thesis?

is shorthand for the conditional

[2] ‘if this (wooden) table were made of ice, then it would look roughly this way.’

But if we then interpret the conditional’s antecedent (‘this table were made of ice’) as necessarily false (since it asserts a metaphysical impossibility), the conditional itself comes out as true. This would mean that the person who has the experience:

IMG PM (there could be a table with thus and such superficial features AND this wooden table when made of ice would look roughly this way)

would in fact be entertaining two true propositions. This suggests that the experience is “veridical” after all, contrary to my suggestion above.

There are at least three possible responses. First, one can simply deny that [1] is logically or semantically equivalent to [2], maintaining that [1] is the proper characterization of the content and that it is straightforwardly false. Resisting equating [1] and [2] may even be supported by the fact that, intuitively, [1] is false and [2] is true. If there is no conditional “hidden” or implicit in [1], then the problem of the vacuous conditional does not arise.

Michael Levin (personal communication) has proposed an interesting alternative strategy. This strategy allows for the presence of a conditional in the content, but, instead of equating [1] with [2], formally analyzes [1] as:

[3] ‘if this (wooden) table were made of ice, there is a way it would look AND any way it would look were it made of ice is identical to that way AND that way is this’

The last ‘this’ refers to the first “existential” portion of the experience’s content (‘there could be a table with q,r,s ’), which is left out of [3], since we are only looking at issues to do with the second half of imaginative contents. [3] emphasizes the fact that, in visualizing a table (or anything), one visualizes it as looking one particular way—from one

specific vantage point—such that each part of the table is determinately to the right or left (or above or below) each other part. Tye notes this feature of mental images, arguing that, unlike discursive descriptions, mental images *invariably* represent a certain single set of determinate spatial relations between any two objects (or two parts of a single object) they represent (1988, p. 515).

So, it is a special feature of visual imagination that objects are represented as looking one particular way; when we don't visualize, but simply think about a table, we don't have to think about it as looking any particular way at all—we need not think of each part of it as being determinately to the left or right of each other part. Thus, despite its complexity, Levin's approach is independently motivated by the fact that visual contents are being described—contents that represent an object as looking one and only one way (i.e. from a certain single vantage point).

Like [1], [3] is false. The first conjunct is trivially true since, like the conditional of [2], its antecedent is false. Yet the second conjunct is false. To see why, note that the triviality of the first conjunct results in its sustaining unrestricted universal generalization. This means that *anything* is a way that the table would look were it made of ice. This is easier to see if we write the first conditional as:

There is a way X such that (if this wooden table were made of ice, it would look way X).

If X is understood as a way that the table could look—a representation of a particular set of perceptible properties by an “image”—anything can be substituted for X while the conditional comes out true (because the antecedent is necessarily false). However, as remarked, it is of the nature of a visualization to represent its object as looking one particular

way. This uniqueness condition is accounted for by the second conjunct: ‘any way it would look had it been made of ice is identical to that way’. This conjunct is false, since, given the unrestricted substitution allowed by the first conjunct, there are (potentially infinitely) many ways the table would look that are not identical to the way the imaginer’s current image represents a table as looking. And if this conjunct is false, the overall content is false (i.e. non-veridical) as well, and we can again say that this is why the imaginer is not imagining what he thinks he is imagining.

A last strategy for dealing with the potentially vacuous conditional implied by [1] is to accept that [2] is the correct characterization of the content, while challenging the traditional truth-functional account of conditionals. That is, one may challenge the idea that the conditionals used in natural language have the same truth values as the material conditional of truth-functional logic. As many before have noted, it is counterintuitive to call vacuously true conditionals “veridical”, especially in the context of mental representation, where we are concerned with the role of conditionals in reasoning and psychological generalizations. A person who believes that *if there are dinosaurs in Brooklyn, then there is life on other planets*, believes something true, on the standard truth-functional account of conditionals, even though it would not be appropriate to infer from the presence of dinosaurs in Brooklyn to the existence of life on other planets. This and related considerations have led many to conclude that the conditionals used in common discourse do not have truth values at all. Rather, they are “devices for letting people know facts about one’s conditional probabilities, which are one’s dispositions to acquire certain beliefs under certain circumstances” (Bennett, 1995, p. 338).⁸⁴ People assent to conditionals with varying degrees

⁸⁴ Bennett is approvingly summarizing Adams (1975). See also Jackson (1987) and Bennett (2003, ch. 7).

of confidence, and in so doing express their views of the probability of the consequent, given the antecedent. On this sort of view, assenting to the conditional ‘if this wooden table were made of ice, then it would look roughly this way,’ is epistemically problematic—for it expresses one’s view that something (the consequent) would likely follow from a state of affairs that is in fact impossible. Clearly, nothing is made *likely* by an impossible state of affairs, since the impossibility itself can never occur. So, believing a content that involves such a conditional results in what we might loosely call a ‘non-veridical’ representation after all; it is a mental representation that harms one’s epistemic standing.⁸⁵

Each of these strategies for dealing with the problem of the vacuously true conditional has its advantages and drawbacks. The first is obviously the simplest, yet it is hard to dispel the feeling that a conditional is “hidden” in [1]. The second admirably respects the presence of a conditional while not requiring a revisionary account of the truth value of conditionals, however one may question whether the content could really have such a complex structure (though the discussion of complexity concerns in Chapter 2 Section 3 may allay such concerns). Also, it is less than intuitive to hold that the above imagining is non-veridical *not* because there is no way that a wooden-ice table would look, but because *any* way is a way it would look. And while the third strategy preserves the presence of a conditional, it is only at the price of a revisionary (if somewhat popular) account of conditionals.

⁸⁵ While it is somewhat controversial to hold that conditionals lack a truth value, Bennett (2003, p. 21) observes that most writers in this area at least accept that conditionals where there is no suitable connection between antecedent and consequent are *false* (his example: “If I ate an egg for breakfast this morning, you at a million eggs”), and not vacuously true (in contrast with the truth-functional analysis of the material conditional). This provides another possible way of dealing with the problem above, if, as intuition would suggest, there cannot be a “suitable connection” between a metaphysically impossible antecedent and a consequent.

For ease of exposition, I will continue to express the contents in the form of [1], though the forms of [2] and [3] should be considered live possibilities as well.

7. Are IMG PM Experiences Reliable Guides to Possibility?

Supposing that all this is right—that Kripke accepts an appearance/reality distinction with imagination, and that his notion of a QIES maps onto my theory of Modal Imagination roughly as I have suggested—a crucial question faces us. Why think that imagination is at all *reliable* as a guide to metaphysical possibility—especially given that we are often wrong about what we are imagining? After all, it is one thing to claim that imagining amounts to entertaining certain kinds of contents, quite another to suggest that, in entertaining these contents, we gain a reliable insight into whether certain things are possible. The reliability question is doubly pressing given that the relevant kind of possibility is metaphysical possibility—which can involve any number of physical impossibilities—and not, say, nomological or epistemic possibility.

In considering the issue of reliability, we can start with a fairly basic determination of possibility: John wants to know whether the envelope he is holding can fit, without bending, through the mail slot at the post office (let's assume he has never put such a letter through the slot, and so memory alone cannot answer his question). On my theory, he would undergo an imaginative experience such as:

IMG PM (there could be an envelope going through a slot⁸⁶, AND this envelope going through the slot at the post office would look roughly this way).

⁸⁶ I am leaving out many additional details (regarding relative sizes, orientations, colors, textures, etc.) that would be represented by the imagistic portion simply for ease of exposition.

If we suppose that John is familiar with how the slot actually looks, and is holding the envelope in question, then, *prima facie*, it seems reasonable to think that his imaginative experience will be veridical; he will be right that the letter can fit through the slot. Not only that, the experience will have served as a guide to determining whether the letter can fit through the slot.

Yet, is it really right to say that the above experience is a *guide* to possibility? Isn't John simply tokening a kind of *belief* that the letter will fit through the slot? After all, both aspects of the content are purely assertoric, and don't involve extracting new information from the environment. What seems to justify the determination of possibility is not the imaginative experience itself, but rather certain past and present acts of perception.

But, what complicates this picture is that, prior to his undergoing in the imaginative experience, John does not have a standing belief about whether it is possible for the envelope to fit. It is not as though he has a *memory* of such a letter going through the slot, and that *this* is what justifies his belief. For, intuitively, if he had such a belief, he would not have to engage in the imaginative experience to begin with. His experience, rather, is comparable to that of someone who, through reflection, becomes surprised at an implication of a standing belief.

Now, one could insist that John has a *latent* belief all along that the letter would fit through the slot, and that his imaginative experience simply consists in that belief becoming occurrent. Yet, this considerably stretches the category of what we normally consider latent beliefs. A moment ago my belief that the street outside my window is not made of sponges was very latent indeed. Yet I did not need to *do anything* to make it occurrent, other than consider the question. By contrast, John needs to do something with the information he has

stored—he must combine disparate elements from memory into a dynamic representation in order to determine whether the letter will fit. This is why it is intuitively coherent to deny that John has a belief concerning whether or not the envelope will fit through the slot prior to his act of imagination, whereas I do have a belief concerning whether or not the street is made of sponges prior to my becoming aware of that belief.⁸⁷

If that is right, then modal imagination should be seen as a faculty that guides our reasoning about possibilities by *enabling the formation of* new commitments as to what is possible. But it is important to recognize that the contents at play in imagination *just are* our commitments to the relevant possibilities; the imaginative experiences are not “mere entertainings” (or “conceivings”) on the basis of which we then adopt a commitment that thus and such is possible. Rather, they *constitute* both the appearance and the commitment⁸⁸ that thus and such is possible. So, it is correct, in one sense, to question whether imaginative experiences are properly called “guides” to possibility, as the experiences simply constitute our commitments about what is possible. Nevertheless, modal imagination, considered more generally as a mode of thought that enables, through an inference-like process, the formation of new commitments about possibility, can still in a sense be thought of as what guides us—or *enables* us—in our determinations of what is possible.

⁸⁷ A closer analogy to the case of imagination would seem to be where one works out the answer to a math problem in one’s head. When trying to calculate 15% of 23, it is not as though arriving at the answer involves awakening a latent belief, though neither is any new information extracted from the environment in answering the problem. What is it, then, that *justifies* our answer in such cases? It must be the reliability (if any) of the process used. Just what that process *is* is famously questioned by Kripke (1982). Intuitively, it seems to me that doing math in one’s head consists in breaking the problem into a sequence of calculations about which one has standing beliefs (e.g. that 15% of 23 = 10% of 23 plus half of 10% of 23, and so on). What justifies the standing beliefs is, perhaps, conformity with the practice of one’s community.

⁸⁸ In Chapter Two I discuss why such “beliefs” are nevertheless different from paradigmatic propositional belief in various ways.

We can now look at the question of the *reliability* of imagination as a guide to possibility with more clarity. Given that imagining just is the forming of commitments (I'll call them "i-beliefs") about what is possible, the question that should concern us is whether these i-beliefs are formed in a way that ensures their reliability. Here I think it is useful to look back to the empirical work discussed in Chapter One on the key role of sensory imagination in memory and in envisioning the future (Addis, et al., 2007; Schacter, et al., 2007). A single "core system" in the brain underlies the human ability to retrieve detailed memories of personal histories, and to plan new courses of action through envisioning them. As Arp (2008) discusses in detail, these representational abilities were presumably evolutionarily advantageous; assuming they arose through natural selection, the *purpose* of imagination is to represent past and future courses of events (and not, for instance, metaphysically possible but physically impossible states of affairs). Also, as Rubin and Greenberg (1998, 2003) discuss, deficits in one's ability to imagine one's past are often correlated in interesting ways with more general visual deficits. As we saw in chapter two, an inability to name by sight objects that one can identify by other means correlates with profound amnesia. This suggests that imagination and vision draw on a common set of neural mechanisms—dissociable from other areas involved in representation—in order to represent. A wide range of other findings support this link. Various visual illusions (e.g. Muller-Lyer illusion, the McCullough Effect) have been found to be mirrored in imagination (Finke, 1989; Finke & Shepard, 1986); performance of visual and imagery tasks tend to interfere with each other in a way that visual and non-imagistic tasks do not (Brooks, 1967);

disruption of activity visual area V5 impedes performance on imagery-related tasks (Kosslyn, et al., 1999).⁸⁹

The close link between vision and imagination suggests some fairly unsurprising constraints on the reliable generation of i-beliefs. The first is a background constraint: an imaginer must have a normally functioning cognitive/visual system for any deliverances of imagination to be considered reliable. The second is a *positive* constraint, detailing the kinds of properties imagination can reliably represent: where the contents of an imaginative experience concern objects and events with whose visual appearance one is familiar (through prior acts of perception, or (perhaps) testimony), and when they involve representation of exclusively visible features of these objects and events (and not, say, microstructure), then, *so far as those properties go*, we can expect such modal imagination to be reliable indicators of what was, or would be the case.

By “visible features” I simply mean the properties typically represented by vision. While drawing a sharp line between visible and non-visible properties is a far from simple task, it is relatively uncontroversial that normal vision represents, at a minimum: shape, color, and some other common macroscopic properties of objects, such as wetness, texture, and rigidity; opinions diverge as to what other properties (natural kinds? emotions?) are *also* represented by vision.⁹⁰ Where imagination represents visible properties in past and

⁸⁹ It is important to note, however, that some evidence (e.g., from brain damage patients) weighs against a complete mirroring of the processes underlying vision and imagery, as some cortically blind patients report intact visual imagery (Bartolomeo, Bachoud-Levi, & Denes, 1997; Behrmann, Moscovitch, & Winocur, 1994; Chatterjee & Southwood, 1995). The point being pursued here is that there is substantial overlap in the two processes—vision and imagination—not that they are exactly the same, or that there can never be partial dissociations.

⁹⁰ Yablo offers an intuitive account of “perceptible properties” along the similar lines. “A property is perceptible,” says Yablo, “iff when an object perceptually appears to have it and doesn’t, we have misperceived. Not all properties figuring in the content of a perceptual state are perceptible in this sense. Our experience may

counterfactual situations, as instantiated in objects with which we are familiar (or simply *some* objects of a kinds with which we are familiar), it can be expected to be reliable.

This suggests a related *negative* constraint: an imaginative act cannot, however implicitly, reliably represent non-actual states of affairs that vision would not reliably represent were they actual. More formally, for any property *P* or relation *R*, modal imagination will only reliably represent that *x* could be *P*, or that *R* could hold between *x* and *y*, if, in suitable viewing conditions where *P*(*x*) and *R*(*x*,*y*), vision would reliably represent *x* as being *P* or *R* holding between *x* and *y*. In short, if vision alone won't tell you the answer in the actual case, imagining will not give you a reliable answer in counterfactual cases.

To see how these constraints play out in specific cases, consider again the post office example, where John undergoes the following experience:

IMG PM (there could be an envelope going through a slot AND this envelope going through the slot at the post office would look roughly this way).

Both components of the content are plausibly reliably formed, since John is visually familiar with such a slot and envelope, his visual/imagery system is functioning normally, and there is nothing represented by the content of his imaginative experience that takes him beyond the negative constraints outlined above (he could tell, by looking, whether or not the envelope was fitting through the slot were he to watch it happening). Now recall again the original ice table example, where the following experience is had:

IMG PM (there could be a table with thus and such superficial features and this wooden table when made of ice would look roughly this way).

represent the table as wooden, but it's not as if our eyes are playing tricks on us if it's well-disguised ice." (2006, fn. 11).

In this case, the possible situation thought to be imagined is not imagined, because (presumably) the second condition is not fulfilled (see again the discussion of Section 7 for alternative interpretations). The particular wooden table would not look any way at all were it made of ice, since things that cannot exist cannot have an appearance. Nor, independently, do we have reason to think that this was a reliable use of modal imagination, since the second half of the content—wherein it is assumed that a certain wood table could have been made of ice—violates the above negative constraint. For it involves representing a certain wooden table as possibly being made entirely of ice, which implicitly requires representing the relation of identity holding between a wooden table and an ice table. Regardless of whether such a relation could hold, it seems clear that *vision* is not a reliable guide to whether it could hold. After all, suppose we saw a wood table being gradually sanded down, with ice replacing the wood at each stage. At the end of the process, we might have a commitment about whether or not the table we ended with was the same table as the one we started with. But it does not seem that vision is what would motivate our commitment; it is not as though we would be in a better position to make the judgment than someone who did not watch the procedure. What motivates it is a commitment—who knows its provenance—concerning criteria for object-sameness across changes in constitution. So, regardless of whether the commitment is reliably formed in some way other, it appears that modal imagination will not itself be a reliable way to form such a commitment.

It appears that the role of sensory imagination as a reliable guide to possibility is quite limited. It tells us very roughly what certain things we are familiar with would look like, or did look like, in relatively pedestrian situations. But there appears no reason to think that, as

a faculty, it underwrites the reliable generation of commitments concerning metaphysical possibilities that are not also physical possibilities, or even concerning a wide array of physical possibilities that involve relationships and properties that are not observable.⁹¹ Of course, it is still not too shabby a cognitive faculty to possess; certainly its loss would be (and is) catastrophic to day to day functioning. It is precisely because we rely so heavily upon imagination to anticipate and plan for future events, and because it serves those purposes so well, that we are inclined to trust its deliverances in less quotidian contexts involving ice tables and substances that are macroscopically like our own natural kinds but have different micro-structures. In short, even when it overreaches, modal imagination invariably involves appearances of possibilities because, in its normal uses, it reliably represents possibilities.

8. *IMG PM and the Mind-Body Problem*

Where does this leave us with Kripke's account? I began by arguing that Kripke uses "imagine" as a success verb, allowing for veridical and non-veridical cases of imagining. I then connected my notion of imaginative content to Kripke's appeal to QIES, showing how in sensory imagination one merely apparent possibility may be mistaken for another genuine one. Further, I have justified the idea that there exist such things as apparent possibilities by

⁹¹ One might object that sensory imagination might nevertheless reliably represent physical impossibilities (but metaphysical possibilities) that would presumably be visible—such as a fork defying gravity and flying up to the ceiling. But, given that gravity (and its absence) is not *visible*, this would overstate the powers of sensory imagination. Some of the intuition that we can correctly imagine a fork defying gravity and flying to the ceiling can be recovered, however, if we emphasize that sensory imagination is a reliable guide to what a fork would look like were it to fly to the ceiling *tout court* (i.e., without comment on whether this resulted from the breaking a law of nature, or a toddler's throw).

Now, perhaps the laws of nature could have been different, and so one *can* imagine a fork defying the law of gravity. The point remains that sensory imagination, *qua* faculty that generates beliefs about possibilities, would not be a reliable guide to such a thing being possible, even if it could *represent* such a possibility. For the belief (or, imaginative "commitment") that a fork could violate the laws of gravity to count as (modal) knowledge, it would have to have some other justificatory basis. Again, this chapter is only investigating the justification that sensory imagination provides for the belief in various possibilities.

showing that some cases of imagination are reliable guides to possibility, while indicating the limited scope of this capacity.

Having developed an account of sensory imagination and its relation to possibility, I'd like now to apply this account to Kripke's argument against mind/brain identities. The schematic theory I have been developing since chapter one has an array of independent motivations; it was not developed with the purpose of explaining away, or justifying intuitions about mind/brain separability. My hope is that whatever implications it has for the mind-body problem will be granted a higher degree of credence than other accounts of the relation between imagination and possibility that are entirely motivated by a desire either to vindicate or embarrass the common intuition that the mind is separable from the brain.

It is in Kripke's argument against mind/brain identity theories that his concern with *apparent* possibilities is most vivid—in particular, he is concerned with the apparent possibility that there could be pain without C-fiber firing⁹² (against type-type identity claims), or that a token pain (which he names 'A') could have existed in the absence of the token brain state (named 'B') with which it is correlated. For, Kripke argues, if 'pain' and 'C-fiber firing' are rigid designators for mental and neural types, and if 'A' and 'B' rigidly name a token pain and brain state respectively, then identity statements between the respective pairs must be necessary if true at all. And yet, Kripke says, "it would seem that it is at least logically possible⁹³ that *B* should have existed without Jones feeling any pain at all, and thus without the presence of *A*" (1980, p. 146). Similarly, he calls the type-identity

⁹² Where "C-fiber firing" is simply a placeholder for whatever neural processes empirical science wishes to identify with pain in humans.

⁹³ Kripke identifies logical possibility and metaphysical possibility.

theorist's commitment that there could not be pain in the absence of C-fiber firing, and vice versa, "surprising and counterintuitive". Anyone who wishes to deny the intuitions that the pain and brain state could come apart in each case "must explain these intuitions away, showing how they are illusory" (p. 148).

Kripke of course recognizes that the identity theorist may wish to adopt Kripke's own strategy to explain away other appearances of contingency—namely, an appeal "illusory" modal appearances resulting from QIESs. But this will be to no avail, Kripke claims, because there is no such illusion present in the pain/C-fiber case. In the heat case, for instance, what was really possible was the subject being in an epistemic situation qualitatively similar to the one he is in the presence of heat—to have sensations *as of* heat—in the presence of something other than molecular motion (and thus other than heat). Similarly for the ice table: it is possible, when viewing an ice table, to be in a situation qualitatively identical to the situation one is in when viewing a certain wood table. But no such confusion seems possible in the case of pain, if we accept with Kripke that there is no difference between having a sensation as of pain (i.e. seeming to be in pain) and simply being in pain.⁹⁴ For, Kripke observes:

To be in the same epistemic situation that would obtain if one had a pain *is* to have a pain; to be in the same epistemic situation that would obtain in the absence of a pain *is* not to have a pain. The apparent contingency of the connection between the mental state and the corresponding brain state thus cannot be explained by some sort of qualitative analogue as in the case of heat (1980, p.152).

So, according to Kripke, when it seems that there could be pain in the absence of this or that brain state, the identity theorist cannot say that we have mistaken this (im)possibility with a *different* genuine possibility of the kind previously adduced—one where there is a mere

⁹⁴ Certainly plenty of people reject this claim, and doing so changes the field of play considerably. For the sake of argument, I am looking at what can be done if the claim is accepted.

sensation *as of* pain but no C-fiber firing. For that, on the current assumptions, is a situation where there is pain indeed; and, if pain can occur in the absence of C-fiber firing (or token pain *A* in the absence of token brain state *B*), then the identity theory fails.

Taking stock briefly, this argument makes clear that Kripke thinks there are such things as modal appearances (which, he grants, *may* be illusory), that these appearances weigh strongly against both type and token identity theories, and that it is not at all clear how these appearances might be misleading one in the pain/brain state case.

Earlier I argued that Kripke's account of QIES maps fairly cleanly onto my account of modal imagination. Now I want to show that, if that assimilation is correct, then there are two reasons to reject Kripke's claim that the identity theorist is in a pinch. First, there is indeed a relevant QIES in the case of pain; second, it is likely that the modal imagining that generates the sense of contingency in the first place is unreliable. I'll explain each in turn.

Up to this point I have only discussed *visual* cases of imagination. However I think the general picture can, with some caution, be extended to other forms of sensory imagination, and even to haptic cases of imagination—in particular, imagining pains. What was the “image” in the case of visualization will constitute what Martin (2002) and others have called the “sensory core” for sensory imagination more generally. That is, we can distinguish on the one hand between the purely sensory-perceptual aspect of the imaginative acts—which, again, may be featured in the imagination of a variety of different token objects and states—and whatever additional content is “conjoined” to that aspect, allowing for a two-part structure in all imaginative content similar to the kind I have detailed for vision.

When theorizing about pain, it is essential to distinguish sharply between pains-as-sensations (where these are correlated with if not identical to neural events of some kind) and

pains-as-objects (where these consist in tissue damage or irritation of various kinds), since both phenomena are routinely called ‘pain’ in daily discourse. That is, it’s natural to say that a person who has lost his arm cannot have a pain (as object) in that arm, yet nevertheless can be “in pain” (as sensation) in virtue of having a sensory experience as of a pain (as object) in that arm. Confusingly, the latter sensory experience—being “in pain”—is typically held to be sufficient for one “having a pain” (though, in the phantom limb case, it is a legitimate question *where* the reified pain is). Note that nothing in this distinction weighs in favor or against the idea that pains are essentially conscious.

The sensory core of an imaginative experience as of a pain in my elbow may be the same as the sensory core as an imaginative experience as of a phantom pain in my elbow: both sensory cores have a certain phenomenological quality and represent the presence of a (possible or past) pain in the elbow (I leave it open whether the phenomenological quality may consist purely in its representational features). However, the larger imaginative projects differ, as one involves imagining a situation where there actually is a pain-as-object in the elbow, while the other involves a situation where there is a hallucination of a pain-as-object.⁹⁵ We might represent the differences in the content of the two imaginative projects in this way:

Imagining a real pain: IMG PM (there could be a pain-as-object in my elbow AND if I were to have a pain-as-object in my elbow it would feel like this)

⁹⁵ If we accept that a person’s seeming to himself to have a pain-as-object is sufficient for his having a pain-as-sensation, then the two imaginative projects both involve imagining situations where there is a pain-as-sensation (even if the pain-as-sensation is not the *object* of either imagining—just as visual experiences are not the objects of visualizations).

Imagining a phantom pain: IMG PM (there could be a pain-as-object in my elbow AND if I were to hallucinate a pain-as-object in my elbow it would feel like this)

According to my explication of the notion of a QIES, one is imagining qualitatively identical epistemic situations whenever one uses the same image in the respective imaginative projects. So, even if I imagine a mere hologram of a wooden table, this will be to imagine an epistemic situation that is “qualitatively identical” to a situation I might imagine with an actual wood table, so long as I use the same image both times. Turning back to pain/brain identities, the identity theorist must deny the possibility of a pain-as-sensation in the absence of any C-fiber firing (or, in the token identity case, for there to be *A* in the absence of *B*). And yet such things seem possible to Kripke and others. It seems one can easily imagine pains-as-sensations in the absence of C-fiber firing, and vice versa.

Assuming, for the moment, that we can indeed imagine sensations, and not merely the things sensations are *of*, the imaginative experience of such an act would be:

IMG PM (there could be a pain-as-object in the elbow AND if I were to have a pain-as-sensation *K*⁹⁶ in the absence of C-fiber firing it would feel like this)

The type identity theorist must hold that this is a non-veridical act of imagination and that the satisfaction conditions are not fulfilled. Why does he think they are unfulfilled? The type identity theorist denies the second conjunct—that there could be a pain-as-sensation *K* in the absence of C-fiber firing—since he thinks pains-as-sensation *K* are identical to C-fiber firing. His reasons for thinking this need not concern us for now—the question is whether he can

⁹⁶ Let ‘pain-as-sensation *K*’ name a pain-as-sensation that represents a pain-as-object in the elbow.

explain away the sense that the above imaginative act is “veridical” in line with Kripke’s more general appeal to QIES.

It appears he can, for all that need be the case is that the same image one uses to (try to) imagine a pain-as-sensation in the absence of C-fiber firing can also be used to imagine some other genuine possibility. If we suppose that pain is identical to C-fiber firing, and that C-fiber firing is not identical to X-fiber firing (where ‘X-fiber, firing’ is a brain state of some other type), then the type identity theorist can hold that the following, involving the same sensory-core “image” as earlier, is a veridical act of modal imagining:

IMG PM (that there could be a pain-as-object in my elbow AND if I were to have a pain-as-sensation *K* in the absence of X-fiber firing it would feel like this)⁹⁷

Further, to imagine this situation, the identity theorist may hold, is to imagine a situation qualitatively identical to a situation where, *per impossible*, there is a pain-as-sensation in the absence of C-fiber firing. Put another way, one might think that an imaginative experience as of pain-as-sensation *K* in the absence C-fibers was veridical because one can use the same image to imagine pain-as-sensation *K* in the absence of X-fibers firing—this is the mitigated sense in which the initial imaginative act was “on to something”.

If that is right, then the case of mind/brain identities can be assimilated to Kripke’s more general paradigm for explaining away appearances of contingency by invoking QIES. Nevertheless, one could insist that this represents a mere stalemate in the debate over mind/brain identities. There will, of course, always be other things one *could* imagine with

⁹⁷ For reasons explained below, I do not hold that this is a reliable act of modal imagination—if our belief that such a thing is possible is reliably formed, it is not due to the role of modal imagination. Nevertheless, all that is required here is that there be a *veridical* case of modal imagination that incorporates the same image, not necessarily a reliable one.

the same image—the question is why think we fell into mistake in this case? If imagining is a generally reliable guide to possibility, maybe the initial imaginative act, where one tries to imagine a pain-as-sensation in the absence of C-fiber firing, is “veridical” after all.

Without appealing to the usual considerations of causal interaction between mind and brain, and the general parsimony of a physicalist world-view, the identity theorist has a ready reply. Recall that, in the case of visualization, our imaginative acts are reliable when they represent things that we would be able to discern through vision alone. In the case of sensory imagination involving pain, we can expect that such imaginative experiences will be reliable in representing what certain pains-as-objects would (or did) feel like, since these are the kinds of properties actual pains represent.

The identity theorist can point out that the use of imagination in question is not a *reliable* use of imagination, because it represents states of affairs to which sensations of pain do not give us reliable access. That is, sensations of pain reliably provide information about the various pains-as-objects we may have—e.g., arthritic pain in one’s knuckles, or a searing headache. But they do not tell us much about *themselves*; and, certainly, they do not tell us about (or represent) the kind of brain state which supports them and which neuroscience may discover them to be identical.⁹⁸ So then, if the impetus for one’s intuition that pain sensations are not identical with brain states lies in an act of sensory imagination, there is good reason to question the reliability of this intuition, even apart from whatever general considerations might weigh against it. That combined with the suggested means of explaining the intuition

⁹⁸ Byrne makes much the same point: “Pains-as-objects are events that we would be expected to have some interesting modal opinions about, as we do about other familiar perceptible things like presidents and tables. *The means by which* we know about pains as objects [i.e. pains-as-sensations] is a different matter entirely; one should not expect us to have interesting modal opinions about *that*” (2007, p. 140).

away is enough to disarm the argument against both kinds of identity theories, absent some alternative suggestion for how the intuition of contingency legitimately arises.

9. Conclusion: an extrapolation of Boyd?

The form of explanation I've just offered can be seen as an expansion of a general strategy for defusing Kripke's anti-materialist argument first offered by Richard Boyd (1992). Boyd argues that Kripke's paradigm for explaining away illusions of contingency—by appealing to alternative objects or kinds that share superficial properties to the things one thought one was imagining—can be extended to the pain/C-fibers case in a way that Kripke overlooks. For one can grant that anything that seems to be a pain *is* a pain, and so give up the possibility of appealing to something that merely seemed like a pain, while nevertheless appealing to a mere appearance attending the *other* side of the identity. Since, like 'water', 'C-fiber firing' refers to a natural kind, it is natural to suppose that there may be other things that bear (mere) superficial resemblances to cases of C-fiber firing. One can then hold that, when it seems one can imagine pain without C-fiber firing, this is because one is imagining pain without imagining the superficially similar *X-fiber* firing; and, when it seems one can imagine C-fiber firing without pain, this is because one is actually imagining *X-fiber* firing without imagining pain.

This appears to allow a means of explaining away the appearance of contingency that conforms to Kripke's general paradigm. Why, then, go in for the intricate apparatus of imagination that I develop above? The answer can be brought out in considering an objection Crispin Wright (2002) levies against Boyd's proposal. Wright argues that a Boyd-like explanation is too strong, since "any apparent contingency whatever could be claimed

with impunity to be a posteriori necessary—since any apparently perfectly lucidly conceived scenario in which it failed would be properly describable merely as one in which it *appeared* to fail” (2002, p. 416). Wright’s worry is that Boyd’s proposal would allow one to claim that all apparent contingencies are illusory, the illusion being due to the fact that something with a similar outward appearance was what one was actually imagining as separated from the thing in question.

Wright overstates the potential problem in this particular passage, however, since if successful Boyd’s strategy is only sufficient for establishing that there always *may* be a necessary relation where there is an appearance of contingency, not that there always *is* one. The more legitimate problem Wright identifies is that Boyd’s suggestion threatens to completely sever the presumed link between imaginability (Wright uses ‘conceivability’) and possibility, by confining imaginings to appearances. “The link between conceivability and possibility may be subtle and qualified,” says Wright, “but it is genuine—and to confine conceivability to appearances would be to sever it altogether” (2002, p. 416). Wright notes that if imagination can only tell us about the appearances of the things we are imagining, many of the key examples motivating the doctrine of a posteriori necessities are called into question. Boyd’s argument threatens to burn the entire modal-epistemological house down just to assimilate the mind/brain case to a set of less controversially mistaken intuitions. Accepting, for the sake of argument, Wright’s claim that the link between conceivability and possibility is “subtle and qualified” but “genuine”, Boyd seems to offer a sledge-hammer where tweezers are needed.

A goal of this section has been to make it intuitively plausible that an argument of Boyd’s general form is correct, by showing how the allegedly mistaken intuitions in question

are likely linked to the kinds of imaginative experiences that are typically used in modal determinations. But by focusing solely on sensory imagination—and making a Boyd-like argument *just with regard to sensory imagination*—I’ve left open a way of circumventing Wright’s “too general” criticism. This is because, for all that has been said, beliefs about metaphysical possibilities (and impossibilities) that are generated without the use of sensory imagination may still more or less reliably track truth.

Consider, for example, the intuition that a certain wood table could not have been made of ice. This doesn’t seem to be connected to, or generated by, any particular act of sensory imagination (what its basis is, and whether it is reliably formed, I simply leave open). The same goes for the intuition that any non-H₂O liquid cannot possibly be water, or that gold could not possibly have atomic number 78; if anything, the use of visual imagery seems to muddle such intuitions. Thus, plenty of our intuitions concerning metaphysical possibilities and impossibilities can still be assumed reliable (if one was previously inclined to assume it), even if sensory imagination is much more limited in its modal deliverances than is often supposed.

So, to sum up, the apparatus of the present theory of imagination is useful when considering Kripke’s anti-materialism because: 1) it provides, through the connection of imagery to QIES, a way of seeing how there can be a relevant QIES in the case of the supposed pain/C-fibers identity, 2) it gives an independent reason to think the particular use of sensory imagination in question is not a reliable means of forming beliefs about metaphysical possibilities, and 3) it allows this intuition to be explained away without all modal intuitions being impugned.

Now, the anti-materialist could circumvent my entire argument simply claiming that *her* intuition that pains can exist in the absence of any token brain state is not grounded in any act of sensory imagination. I have tried to make it plausible that imagery is involved in generating such intuitions, but I do not take myself to have given an air-tight argument to that effect—nor do I think one could be given (it’s just too easy for someone to say “*that’s* not why I believe that”). Many of the most sophisticated defenders (e.g., Chalmers & Jackson, 2001; Levine, 2001, Ch. 3) of the idea that there is a significant metaphysical or epistemological gap between mind and brain base their claim not (directly) on an insight arrived at through sensory imagination, but on the putative fact that the “higher-level” facts about consciousness cannot be derived or inferred from “lower-level” physical or functional facts (as though they were sitting at home one day, cranking out such derivations across a variety of domains, and suddenly realized that the derivations just didn’t work out in the mind/brain case).

Like most people, I sense a disconnect when I try to imagine a conscious visual experience on the one hand, and *any* kind of physical state to which it might allegedly be identical on the other. But if I force myself to consider the question of the identity while using no imagery at all, the feeling of an unbridgeable gulf wanes—I begin to think that I better just accept whatever the scientists say. Is this a peculiarity of my own philosophically twisted imagination? Maybe.

But maybe not. So many of the metaphors used to evoke the irreducible character of mind involve light: the mind is a “mysterious flame” (McGinn, 2000), an “inner glow”. Where there is a lack of consciousness, “all is dark inside” (Chalmers, 1996, p. 96). Might this be because we typically use visual images—images that are invariably of *well-lit*

environments—to think about (i.e., to *imagine*) experiences? But, if visual images cannot reliably be used to form beliefs about things that cannot be seen, and if visual experiences themselves cannot be seen, *and* if the mind/brain relation does not seem especially mysterious when one does not form mental images while considering it, then at least some of common perplexity surrounding the mind/brain relation is ill-founded.

I think all three antecedents are true. I argued for the first, and the second is relatively uncontroversial.⁹⁹ As for the third—that is between you and your intuitions.

⁹⁹ An exception might seem to be Higher Order Perception (HOP) views of consciousness—however it is unlikely that a HOP theorist would hold that (what we normally call) mental images represent mental states. Rather, the HOP theorist holds that a mental image is conscious when a higher-order (quasi) perceptual state is representing *it*. On such a theory, we would have reason to think that the higher order states reliably represent mental states of various kinds, but not that mental images themselves—the things we typically use to imagine worldly objects—reliably represent mental states.

CHAPTER 4

Propositional Imagination: Imagining as Believing

1. Introduction

The first three chapters of this dissertation have examined sensory imagination, which I defined as all thought involving mental imagery. I want now to switch gears and discuss the sister phenomenon of propositional imagination. Unlike sensory imagination, which includes simple acts of remembering, propositional imagination is typically linked to the notions of pretense and supposition. Propositional imagination—so named because it occurs when someone imagines *that* *p*, where *p* is a proposition (instead of simply (sensorily) imagining *X*, where *X* is an object)—can involve imagery, but need not; this much is in keeping with common usage (T. Gendler, 2005; Nichols, 2006b; Walton, 1990). Since I define non-propositional imagining as all thought involving imagery, this means that *some* propositional imagining may involve non-propositional imagining. I explain how this can be so at the end of this chapter, in section nine. Until then, I will focus on cases of propositional imagining that do not involve imagery, in order to avoid confusion.

A view has taken hold in recent years according to which propositional imagining, and hypothetical reasoning more generally, involves a distinctive cognitive “attitude”. The attitude goes by various names—‘imagination’, ‘pretense’, ‘supposition’—but the core idea is that, just as belief and desire constitute two different attitudes one can take toward a proposition, so too is there an “imaginative” attitude that one can take toward a proposition, accounting for thoughts that are functionally quite different from any beliefs or desires one might entertain.

To hold that people propositionally imagine is, minimally, to hold there are true statements of the form “*X* imagines that *p*,” where *X* is a thinking subject and *p* is a proposition. I will take the claim that there is a distinctive imaginative *attitude* to be considerably stronger, as it is a claim about the *cognitive architecture* that underlies the ability to propositionally imagine.¹⁰⁰

Nichols and Stich (2000; 2003 Ch. 2) develop one of the most influential cognitivist accounts of this imaginative attitude as a way of explaining a range of empirical data concerning human pretense. Their theory seeks to formalize two common claims about propositional imagination: the first is that it does indeed involve a cognitive attitude distinct from both belief and desire; the second is that this attitude bears important functional similarities to belief. These two ideas are packaged together in what Nichols (2004, 2006b) calls the “single code hypothesis,” according to which “psychological mechanisms that can take input from both imagining and from believing will process imagination-based inputs (‘pretense representations’) and isomorphic beliefs in much the same way” (2006b, p. 459).¹⁰¹ For implicit in this hypothesis is the (substantive) claim that imagination and belief

¹⁰⁰ One might dispute whether calling *X* a “distinctive” propositional attitude commits one to any view concerning cognitive architecture (e.g., one might grant that *X* is a distinctive propositional attitude but hold that, at a more basic level, *X*ing reduces to having certain kinds of beliefs and desires). For simplicity, I will assume that to hold that *X* is a distinctive propositional attitude (from belief and desire) is to hold that *X*ing requires entertaining mental states above and beyond collections of beliefs and desires, exploiting cognitive mechanisms above and beyond those that underlie belief and desire. The same assumption guides the work I will be discussing (e.g., Currie & Ravenscroft, 2002; Nichols & Stich, 2000). See also Weinberg and Meskin (2006a, 2006b) and Schroeder and Matheson (2006).

¹⁰¹ To call the account “single code” is slightly obscurantist, as the most salient feature of the view is that it posits a propositional attitude *in addition to* belief and desire (and correspondingly distinct cognitive mechanisms). Obviously, views (such as the one I will argue for) that do not posit this distinctive attitude can equally well posit representations “in the same code.” Both Weinberg & Meskin (2006a) and Schroeder & Matheson choose preferable terminology in positing essentially the same kinds of mechanisms for propositional imagination. For Weinberg & Meskin, single code theories of imagination accept what they call “Minimal Nondoxastic Cognitivism” (MDC) about imagination, while Schroeder & Matheson (2006) call single code theories “Distinct Cognitive Attitude” (DCA) theories. Thus, in arguing against N&S’s single code theory, I am also challenging Weinberg & Meskin’s claim that a satisfactory theory of imagination requires accepting MDC,

are indeed distinct cognitive attitudes, and that the imaginative attitude is importantly belief-like (its similarity to belief consisting in some of the same mechanisms—e.g. “inference mechanisms”—processing input from both belief and imagination in “much the same way”).

While there have been many proposed amendments and buttresses to Nichols and Stich’s specific development of the single code hypothesis (e.g., Carruthers, 2006; Currie & Ravenscroft, 2002; Tamar Szabo Gendler, 2003; Weinberg & Meskin, 2006a) the two core ideas behind the single code hypothesis are widely accepted (see, e.g., Carruthers, 2006, pp. 89-91; Currie & Ravenscroft, 2002, Ch. 2; Tamar Szabo Gendler, 2006, pp. 183-185; Schroeder & Matheson, 2006; Weinberg & Meskin, 2006b, pp. 178-180).¹⁰² Noting this “remarkable convergence of opinion”, Schroeder and Matheson offer their 2006 as “a final seal of approval” upon the idea that pretense involves the activity of a distinct belief-like cognitive attitude (p. 20), while Nichols remarks that the single code hypothesis is “embraced by most contemporary cognitivist accounts of pretense and imagination” and is “perhaps, the most productive idea about the propositional imagination that anyone has ever had” (2004, p. 129, 2006b, p. 459).

This chapter diverges from the convergence. I argue that the data on pretense that Nichols and Stich, and many others, take to strongly support the single code hypothesis can all be explained without the introduction of a distinctive cognitive attitude (a “DCA”, for short), and without attributing to pretenders any unduly complex beliefs or desires. Moreover, I argue that childhood pretense can be explained while attributing no beliefs or desires to pretenders over and above those that proponents of the single hypothesis must also

and Schroeder & Matheson's claim that DCA theories are the only satisfactory theories of propositional imagination.

¹⁰² For early precursors, see Leslie (1987) and Perner (1991, Ch. 3).

attribute as part of their own accounts. This makes my argument in part an appeal to parsimony.

However, single code theorists usually do not hang their hat on explanations of pretense alone; there is a range of other phenomena that the theory purports to explain as well (see, e.g., Currie & Ravenscroft, 2002, Ch. 7-8; Nichols, 2004; Weinberg & Meskin, 2006a). Nevertheless, pretend play creates probably the clearest case in favor of the single code hypothesis. Indeed, Nichols and Stich (2003) begin their book-length treatment of mindreading¹⁰³—which they argue requires the same distinctive imaginative attitude—by an appeal to mechanisms needed for pretense. As they explain, “thinking about pretense and the cognitive mechanisms that underlie it played a quite pivotal role [in shaping their thinking on mindreading], because it convinced us that the basic cognitive architecture [involving beliefs and desires] needed to be supplemented” (2003, p. 16). So, for Nichols and Stich, it is pretense that most clearly requires appeal to a DCA; once such a DCA is in play, it can potentially be called upon to explain other things as well.

My strategy, then, is to attack the single code hypothesis by questioning its greatest supposed virtue, as expounded by some of its most able defenders.¹⁰⁴ That said, there are two other features of imagination and pretense that, while not explicitly addressed by Nichols and Stich (2000), often motivate other discussions (e.g., Currie & Ravenscroft, 2002, Ch. 7-8; Nichols, 2004; Nichols & Stich, 2003, Ch. 3; Weinberg & Meskin, 2006b), and about which I shall have something to say. These are the tendency of (mere) imaginings to give rise to affective and emotional responses in imaginers, and the co-presence of deficits in both

¹⁰³ ‘Mindreading’ refers to the ability to predict and explain of the behavior of others by appeal to folk-psychological states such as beliefs and desires.

¹⁰⁴ Tamar Gendler calls N&S's theory “the most detailed theoretical discussion of pretense to date” (2006, p. 184).

pretend play and mindreading in children with autism. After presenting my alternative account of pretense, I argue that these data ultimately weigh against the hypothesis as developed by both Nichols and Stich (2000) and Currie and Ravenscroft (2002) (from here on I will refer to the Nichols and Stich of their (2000) as ‘N&S’, and to the Currie and Ravenscroft of their (2002) as ‘C&R’).

As a last opening remark, it should be noted that while the issues surrounding imagination and pretense connect in important ways with the well-known simulation vs. theory-theory debate over mindreading, the single code hypothesis itself “is shared by both simulation theorists [such as Gordon and Barker (1994), Currie (1995), and Harris (2000)] and their foes [Leslie (1987) and Nichols and Stich (2000)]” (Nichols, 2006b, p. 459). Indeed, an important battle between the two camps is over which can provide the most plausible cognitive-architectural implementation of the single code hypothesis and, by extension, the best explanation of pretense behavior. So, the argument against the single code hypothesis does not directly engage the argument between simulationists and theory-theorists; rather, it attacks what is usually common ground between the two.¹⁰⁵

My view, in a nutshell, is that imagining that *p* amounts to forming beliefs about what would be the case if *p*, based upon retrieved beliefs about relevant conditionals and generalizations; and, pretending that *p* is acting in accordance with such beliefs (though both these characterizations are omitting some complexities I will come to later).¹⁰⁶ None of this,

¹⁰⁵ That said, there may be some simulationist or theory-theory accounts of mindreading that are not single code accounts. It is difficult to say what is consistent (or not) with either simulation or theory-theories *in general* since instances of such theories differ in key respects, and the very question of whether there is a substantive difference between the two kinds of theory has itself been a matter of considerable dispute (see e.g., Heal (1994, 1998), Stich and Nichols (1997, 1998) and Davies & Stone (2001)).

¹⁰⁶ I will not offer any particular account of belief. The standard, if unilluminating, characterizations will have to do: to believe something is to take it to be the case, or to accept it as true; beliefs are functionally distinguished by their privileged role in guiding action and practical reasoning. I intend these not as

I argue, requires positing a distinctive cognitive attitude; call this view a *single attitude* account of pretense and imagination (though some desire is needed, too). The view also happens to posit representations that are all “in the same code.”

In the next section I lay out the empirical data concerning pretense that N&S (I think correctly) argue that a theory of pretense should help explain. Section Three looks in detail at N&S’s explanation of these data, and Section Four develops my own more parsimonious alternative. Section Five looks at Currie and Ravenscroft’s (2002) single-code account, arguing that it is no better motivated than N&S’s. Sections Six and Seven look at the phenomena of mindreading deficits in autism, while Section Eight addresses the issue of affective responses to the imagined. I argue that, contrary to the claims of single code theorists, these phenomena weigh against single code theories once there is an alternative account of pretense available. Section Seven also develops an independent problem for N&S’s account, through an argument that pretense requires *disbelief* in some premise relevant to the pretense. Finally, in Section Nine, I show how the conception of propositional imagination offered here meshes with the account of sensory imagination developed in the first three chapters.

2. The Data on Pretense

A commonly cited example in discussions of pretense, including that of N&S, comes from Leslie (1994), who reports on an experiment where children have a pretend tea party. It can be used to bring out the key features of pretense that a theory should help explain. Here is the passage from Leslie cited by N&S:

philosophically adequate characterizations of belief, but rather as a mere indication that I am not working with any special or controversial conception of belief.

The child is encouraged to “fill” two toy cups with “juice” or “tea” or whatever the child designated the pretend contents of the bottle to be. The experimenter then says, “Watch this!”, picks up one of the cups, turns it upside down, shakes it for a second, then replaces it alongside the other cup. The child is then asked to point at the “full cup” and at the “empty cup” (both cups are, of course, really empty throughout)” (Leslie, 1994, p. 223, quoted in Nichols and Stich, 2000, p. 117).

2.1. Quarantining

Ten out of ten two-year-olds in Leslie's experiment identified the cup that had been turned upside down as the “empty cup”, and the one that had not been overturned as the “full cup”.¹⁰⁷ One of the main questions motivating pretense research since Leslie's (1987) seminal paper concerns the issue of cognitive “quarantine”: how do children (and adults, for that matter) manage to keep what they believe separate from what they are merely pretending?¹⁰⁸ How does the child, who really believes both cups to be empty, manage to correctly “keep track” of the fact that one of the cups is “full” in the pretense?¹⁰⁹ How does she do so without falling into a kind of representational “chaos” or “abuse” (Leslie, 1987), whereby the cup is represented as both full and empty? From the perspective of cognitive architecture, what sort of mechanisms and mental states make it possible for children to give the “correct” answers to the experimenters, while avoiding representational “chaos”?

¹⁰⁷ The children also easily answered similar questions regarding the pretense that required a verbal response of something other than merely “Yes” or “No”.

¹⁰⁸ Leslie introduced the notion of quarantining in the following passage: “Pretense affects the normal reference, truth, and existence relations of the representations it uses. These relations become highly deviant. Any primary representational system affected would quickly be undermined by arbitrary meaning changes. To prevent this, pretend representations must somehow be marked off, or ‘quarantined’, from primary representations” (1987, p. 415). The hidden premise here—the one that launched a thousand papers—is that, if one is committed to a representational theory of mind, then one must explain pretense behavior by appeal to cognitive agents’ representing that the pretend situation is the case (e.g. that the telephone *is* a banana). For such a “pretense representation” typically “contradicts the primary representation” (*ibid.*). It is precisely this assumption—that pretense requires entertaining representations that are in some sense contradictory—that I challenge in Section Four.

¹⁰⁹ Allowing for what Weinberg and Meskin call “representational distinguishability” (2006a, p. 223).

A related phenomenon this example nicely reveals is that one can pretend (or imagine) that p while believing that not p (the child pretends that a cup is full, while believing it not to be full); and, one can also pretend (or imagine) that p while believing that p ¹¹⁰ (the child both pretends that the overturned cup is empty and believes it to be empty). Any theory of pretense must explain how this can be so (Nichols, 2006a, pp. 6-7). I will group this feature of imagination together with the question of how pretenders distinguish what they believe from what they imagine under the rubric of *quarantining*.

2.2. *Inference*

Another important group of features in pretense can be placed under the rubric of *inference*. Nichols and Stich note that when pretenders elaborate the details of a pretense, they very often do so through a series of inferences that that mirrors fairly closely the beliefs that would be formed were the pretense real. (Nichols (2006a) calls this phenomenon “inferential orderliness,” whereas it is simply called “inferential elaboration” in Nichols & Stich (2000)). In the tea party example, when one of the cups is overturned, the children infer that it has become “empty”, just as they would come to believe it was empty had it actually been filled and then overturned (or if they had simply been told that a full cup, somewhere, was actually overturned). Currie and Ravenscroft discuss the same phenomenon, suggesting that propositional imagination is “belief-like” in that it “preserves the inferential patterns of belief” (2002, p. 12). This phenomenon can be grouped in combination with what Nichols and Stich separately identify as the ability of pretenders to engage in *behavior* “appropriate” to the pretense (2000, p. 119-120). Those engaging in

¹¹⁰ I argue in Section Five that this latter claim requires qualification.

pretense not only make inferences that mirror the beliefs they would form were the pretense real, they also—within certain limits¹¹¹—behave in a way that more or less mirrors the way they would behave were the pretense real. A person pretending that a banana is a telephone will often speak into one end of the banana, just as she would if the banana were actually a telephone.

Nevertheless, as Nichols and Stich also stress, such “inferential orderliness” is only a norm; in many cases things are inferred or imagined in an act of pretense that one would not normally come to believe or act out in reality. If one is told to imagine that Bob was in New York yesterday and London today, one will typically imaginatively infer that he traveled to London by plane (as one would no doubt come to believe were one simply told this information). But there may also be cases where, for whatever reason, one fills out the imaginative scenario by supposing that Bob made the journey via teleportation, or by flapping his arms. So, there is a latitude in imaginative inference—a possibility of divergence from what one would believe were the situation real—that also must be accounted for.

Under the general rubric of inference, then, there are three related questions that a theory of imagination would ideally shed light on: why do the inferential patterns of imagination usually mirror those of belief; how do these inferences connect with and guide a pretender’s outward behavior; and, how can inference and pretend behavior sometimes radically diverge from what we would come to believe or do were the pretend situation real?

¹¹¹ As Nichols and Stich note, most children making “mud pies” do not go so far as to eat them—though they may well cut the pies into slices and put the slices on plates.

2.3 *Conceptual naïveté*

A last important aspect of pretense highlighted by the tea party example is that the ability to engage in pretense develops early in childhood. By two years of age, most children freely and spontaneously engage in group and solitary pretend play (Leslie, 1987). This suggests that one need not have mastered very sophisticated psychological concepts, or be able to process complex inference patterns in order to pretend; whatever concepts and abilities it requires should be ones we can plausibly assume two-year-olds to have.

3. *Nichols and Stich's Possible Worlds Box*

Nichols and Stich's project is to shed light on as many of these features of pretense as they can, while working within the basic framework of "cognitivist" or "representationalist" propositional attitude psychology. They note two key assumptions (both of which I accept) of this common framework. The first is that there are (at least) two very different kinds of mental states—beliefs and desires—that are distinguished functionally, in terms of their distinctive patterns of interaction with other components of the mind. It is typical in diagrams that map out basic cognitive architecture to assign a "box" for each type of state—the Belief Box and Desire Box, respectively. The various boxes are meant to signify that the representations grouped there bear important functional (and not necessarily neurological) similarities. Depending on the details of one's theory of cognitive architecture, these "boxes" have a range of connections with other mechanisms, and with inputs and outputs to the cognitive system.

The second assumption is that beliefs and desires, and whatever other propositional attitudes there may be, involve a relation between an "attitude" and a proposition. Thus, in

describing a thought, we can distinguish between the attitude (e.g., of belief or desire) taken toward a content, and the content itself. Two thoughts can then differ in terms of attitude, content, or both (e.g., the belief *that it is raining* shares content with but differs in attitude from the desire *that it is raining*). Moreover, it is perfectly coherent to believe that *p* while desiring that *not-p*. But it is generally considered irrational to believe that *p* and *not-p*, or desire that *p* and *not-p*.

Nichols and Stich's distinctive claim is that imagining is a propositional attitude *in addition to* belief and desire, which warrants its very own "box" (they call it the "Possible Worlds Box" (PWB); Nichols (2004) later dubs it the "pretense box"). They argue that positing the PWB provides the best means of explaining the various facts about pretense vetted above.

Let us use 'imagination' for the attitude taken toward a proposition in virtue of its being in one's PWB (though 'supposition' would work equally well; note that there is no requirement on their view that mental imagery is involved in—or absent from—pretense). I will understand the relation between imagination and pretense to be as follows (and which I take to be more or less customary): pretense always involves some outward behavior, and is typically carried out in virtue of the various things that the pretender imagines. One can, of course, simply imagine various things without engaging in outward pretend behavior; the point is that when one does engage in pretend behavior, it is guided by what one is imagining. Theories of propositional imagination, then, specify what goes on *mentally* during a pretense.

Looking first at the issue of quarantining, N&S suggest that the PWB accounts for the cognitive system's ability to keep propositions relevant to one's imaginative project separate from those towards which one has the attitude of belief or desire. Of course, the possible world "box" is simply shorthand for a range of functional differences possessed by certain representations in the mind; so, these functional differences in the way the representations are treated ultimately explain how confusion is avoided and accurate reports are given concerning what is believed versus what is imagined.

Moreover, the existence of a distinctive imaginative attitude that one can take toward the propositions guiding pretense helps explain how one can believe that *p* while imagining that not *p*, or imagine that *p* while believing that *p*. For, provided that imagining is, like desire, its own cognitive attitude, there is no obvious epistemic difficulty—no threat of "inferential chaos"—presented by the fact that one imagines that *p* while believing that not-*p*, just as there are no such problems inherent in believing that *p* while desiring that not-*p*.

Turning to the issues under the rubric of inference, N&S hold that it is in the PWB that the inferences and elaborations relevant to guiding a pretense take place. The tea party pretense begins "when a representation with the content *We are going to have a tea party* is placed in the PWB" (p. 122). What happens next is:

The cognitive system starts to fill the PWB with an increasingly detailed description of what the world would be like if the initiating pretense were true...in Leslie's tea party scenario, at the point in the pretense where *Alan has just turned the green cup upside down* has been added to the PWB, the child's cognitive system has to arrange to get *The green cup is empty* in there too. (p. 122).

This new representation will arrive via a process of inference that goes on in the PWB; Nichols and Stich posit that such inferences occur through the work of the same “inference mechanism” that guides the formation and revising of real beliefs. They propose that it is able to work on representations in the PWB in the same way it works on representations in the Belief Box because representations in the PWB are “in the same code” as those in the Belief Box (p. 122-125).¹¹²

Yet, as they are quick to point out, the inference-process will not get very far if the only representation in the PWB is the one that initiates the pretense. There are only so many things that can be inferred from *We are going to have a tea party* alone. Some further commitments concerning what typically goes on at tea parties are needed for the pretense to develop. Nichols and Stich suggest that this extra information has an obvious source: the Belief Box. Along with the pretense-initiating premise, “the cognitive system puts the entire contents of the Belief Box into the Possible World Box” (p. 123). This way, a rich set of inferences can be drawn from the pretense-initiating premise—importantly, it will be a set of inferences that closely mirrors those that would be drawn were we to believe the initial premise, since the very same set of beliefs serve as additional premises.

Of course, some of the beliefs in the Belief Box will conflict with the initial premise (such as, in the above example, the belief that we are *not* having a tea party), potentially leading to “inferential chaos”. They therefore posit an additional cognitive mechanism, the “Updater”, which “goes through the representations in the PWB, eliminating or changing those that are incompatible with the pretense premise” (p. 124). While they note that it is hard to understand how exactly the Updater would accomplish its task, they find it

¹¹² Nichols and Stich hold that representations are in the same code if they have the same logical form, and their representational properties “are *determined in the same way*” (2000, p. 125).

unproblematic to posit it since the same mechanism is needed to explain how we revise and change beliefs.

Concerning the various inferences and behaviors during pretense that do not follow belief-like inference patterns, Nichols and Stich admit they “don’t have a detailed account of the cognitive mechanisms” that account for them (p. 127). They simply posit an additional component—the “Script Elaborator”—whose job it is to “fill in those details of a premise that can’t be inferred from the pretense premise, the (filtered) contents of the Belief Box and the pretender’s knowledge of what has happened earlier on in the pretense” (*ibid.*). As they recognize, positing such a mechanism does little more than provide space for an explanation to come; they posit it mainly to emphasize that such a mechanism is indeed needed, and to underscore “how little we know about how this component of the mind works” (p. 144).

The last issue under the rubric of inference to be discussed is how N&S connect the activity in the PWB to the actual behavior of pretenders. According to N&S, “Pretenders behave in the way they do because they *want to behave in a way that is similar to the way some character or object behaves in the possible world whose description is contained in the Possible World Box*” (p. 128, emphasis in original). While this sounds like a rather complex desire to have, they quickly provide a simpler gloss: “To pretend that p is (at least to a rough first approximation) to behave in a way that is similar to the way one would (or might) behave if p were the case. Thus, a person who wants to pretend that p wants to behave more or less as he would if p were the case” (p. 128).¹¹³ Note that having this simpler desire—“to

¹¹³ Since Nichols and Stich never refine their definition of pretense, it is worth noting that pretending that p is not acting as one would if p, but rather acting as one *believes* one would if p (though, for reasons given in Section Five, this is still not a sufficient criterion for pretense). After all, I may unwittingly have a neurological tic that causes me to go into paralysis whenever I attend a tea party. But, even if that's true, I need not behave as though I am in paralysis (i.e., as I actually would at a tea party) in order to pretend that I am at a tea party. (Thanks to Michael Levin for the example)

behave more or less as [one] would if p ”—is not at all the same as having the desire to act in accordance with descriptions in a PWB; for one thing, it involves mastery of far simpler concepts (I will indicate why I think this matters for their account later).

Of course, the desire to behave as one would if p —I will leave out the redundant “were the case” from here onward—can only be fulfilled if one has some beliefs about how one would behave if p . “The obvious source for this information” they tell us, “is the possible world description unfolding in the PWB” (p. 128). Yet, since the PWB is distinct from the Belief Box:

We must assume that the contents of the former are accessible to the latter. More specifically...we assume that as a possible world description is unfolding in the PWB, the pretender comes to have beliefs of the form: If it were the case that p , then it would (or might) be the case that $q_1 \& q_2 \& q_3 \dots \& q_n$, where p is the pretense premise and $q_1 \& q_2 \& q_3 \dots \& q_n$ are the representations in the PWB (p. 128).

Thus, pretense behavior occurs when the pretender tokens a desire to behave as she would if p , and, by accessing the description “unfolding” in the PWB, is able to form beliefs about how she would act if p .

So then, we have a sketch of how N&S account for the phenomena of cognitive quarantining and inferential orderliness. As for the early emergence of pretense behavior, they believe this fits naturally with their account since “on [their] theory, the pretense could proceed perfectly well even if the subject did not have the *concept* of pretense” (p. 137-138). Unlike other views (e.g., Leslie (1987, 1994)) that require pretenders to have beliefs of the form, “I am pretending that such and such,” N&S suggest that the *attitude* of imagination can be taken toward a content without one having the concept IMAGINE (or PRETEND), just as

Note well, however, that the *desire* needed to motivate a pretense that p is still simply a desire to act as one would if p . For one will naturally draw on one's beliefs about how one would act if p when trying to satisfy this desire. Thus, you need not desire to act as you *believe* you would act if p (and hence need not exercise the concept BELIEVE), in order to possess the mental states required for the pretense.

one can take the attitude of desire toward a proposition without having the concept DESIRE (recall that, on their account, the desire leading to an instance of pretending that *p* is a desire to act as one would if *p*—a desire one could presumably have while lacking the concept PRETEND).

4. Imagining in the Belief Box – a single attitude account

I will now offer a more parsimonious proposal for making sense of the same data. If successful, it weighs strongly against any theory of pretense that appeals to a distinctive cognitive attitude.¹¹⁴

*4.1. Imagining that *p* without entertaining the proposition that *p**

Two key claims of my account of pretense are, first, that imagining (or pretending) that *p* does not require entertaining the proposition that *p*; and, second, one can explain the behavior of pretenders, and the thought-processes and cognitive abilities of imaginers, by appeal to beliefs and desires that single code theorists must also allow pretenders. As we will see, these two claims end up supporting each other.

The claim that imagining or pretending that *p* does not require entertaining the proposition that *p* may strike some as counterintuitive. It is hard to pinpoint the exact reason for this intuitive tension. Certainly we cannot infer that pretending or imagining that *p* requires entertaining the proposition that *p* directly from the fact that there are true statements of the form, ‘X pretended that *p*,’ and ‘Y imagined that *p*.’ After all, there are plenty of verbs

¹¹⁴ It also weighs against Leslie's (1987, 1994) proposal that pretense requires a special “decoupling” mechanism—one that distinguishes the representations involved in pretense from the “primary representations” that underlie belief.

followed by that-clauses that do not transfer to corresponding cognitive “boxes” containing the propositions in question. For instance, Bob can insinuate *that the governor is a crook* without our needing to posit an “insinuation box”, wherein resides the proposition ‘the governor is a crook.’ Nor does the fact that one can insinuate (or implicate, or simply say) *that the governor is a crook* while believing that the governor is *not* a crook, and also while believing that the governor *is* a crook, obviously stand us in need of a special cognitive quarantine area for the things we insinuate (or say). Bob might insinuate that *p* while believing that *not-p* because he wants you to believe that *p*; this involves his having a desire with the content *that you believe that p*. But that does not amount to entertaining the proposition that *p* (even if it involves thinking about many of the same objects and situations that someone who entertains the proposition that *p* thinks about). Why, then, is the situation different for ‘imagines’?

One might think that the term ‘imagines’ transfers to a propositional attitude because, like ‘believes’ and ‘desires’, it seems to create intensional (or “referentially opaque”) contexts when followed by a that-clause.¹¹⁵ Intuition (perhaps) has it that someone can imagine *that Mark Twain is piloting a steamboat* without imagining *that Samuel Clemens is piloting a steamboat*, just as it is often held that Mark Twain piloted steamboats without believing that Samuel Clemens piloted steamboats.¹¹⁶ But there is an equally plausible alternative explanation for the intensionality of ‘imagines’, which is that propositional imagining *reduces to*, insofar as it *consists in*, believing and/or desiring certain things. If that

¹¹⁵ Intensional (or “referentially opaque”) contexts occur in sentences when co-referential terms cannot be substituted while necessarily preserving the truth value of the sentence.

¹¹⁶ As emphasized in chapters one through three, *non-propositional* uses of ‘imagines’ do not seem to create intensional contexts. Intuitively, anyone who imagines Mark Twain (i.e., forms a visual image of him) also imagines Samuel Clemens (and the author of *Huck Finn*, and so on), whether he believes he does or not. Again, this chapter only concerns propositional imagination.

were right (and I will argue that it is), then instances of ‘X imagines that y’ would *inherit* their intensionality from their relationship to these genuine propositional attitudes. For now, this possibility is enough to establish that we cannot move directly from the intensionality of ‘imagines that’ to the conclusion that imagining that x requires entertaining the proposition that x , or to the existence of a cognitive “box” containing x .

Fortunately, few if any single code theorists rest their case solely on these linguistic features (if they did, they would risk arriving at cognitive architecture through conceptual analysis alone).¹¹⁷ Examples like Leslie’s tea party are what really set the view in motion. There were three main facts about the tea party that call for explanation: the phenomenon of “quarantining”, the belief-like inferential and behavioral orderliness of the affair (which I grouped under the rubric of ‘inference’), and the conceptual naïveté of its participants. I will begin my own account by addressing the issues surrounding inference, which will reveal my approach to quarantining and conceptual naïveté.

4.2. *What a tea party requires*

For now, let us be agnostic as to whether the children in Leslie’s experiment must entertain the proposition that the green cup is empty in order to imagine that the green cup is empty. Begin simply with the uncontested data: the child (correctly) answers the experimenter’s question by identifying the cup that was turned over as the one that is empty,

¹¹⁷ See, however, Leslie (1987, 1994), who argues that “the fundamental forms of pretense reflect the semantic phenomena of *opacity*,” which he glosses as, “the ‘suspension’ of the semantic relations of reference, truth and existence that occurs when a representation is placed in an intentional (sic) context” (1994, p. 215). Leslie argues that the three forms of pretense match up to the three main opacity phenomena in a way that requires explanation by a theory of pretense. But, as Nichols and Stich (2000, p. 140-141) also point out, the same opacity phenomena attend reports of other propositional attitudes as well, such as belief and desire. So it is unclear why Leslie thinks that opacity is a special feature of “decoupled” pretense representations, and not of the “primary” representations from which they are (on his view) generated (1994, p. 217).

and the one that was not turned over as the one that is full, while believing all along that both are really empty. What sort of beliefs, desires, and perceptual experiences must the child have to accomplish this?

I agree with N&S that the child must have a desire to engage in the pretense—she must “want to behave more or less as [s]he would if *p*” (where *p* is ‘we are having a tea party’). She must also have some beliefs about what goes on at tea parties. These, too, Nichols and Stich allow the child must have; for these are the beliefs that, once copied into the PWB, enable inference to “fill out” the PWB description in relevant ways (Nichols and Stich call such clusters of beliefs “scripts” or “paradigms” that detail “the way in which certain situations typically unfold” (p. 126)). And, of course, I agree that the child must be able to see (or otherwise perceive) what is actually going on (are you handing her a cup, tipping a kettle into a pouring position, offering her a cookie-sized object?).

My difference with N&S is that, unlike them, I think that, suitably combined, such beliefs, desires, and percepts are sufficient to explain the pretense behavior. To show how this can be so, I will begin by simply listing, in roughly chronological order, one such collection of beliefs, desires, and percepts that the child might have during the pretense. I will try to indicate the grounds for the inferences that take place; most of them are not strict logical inferences, but rather have an evidential basis of some kind.¹¹⁸ Yet these are the same kinds of inferences that N&S posit take place in the PWB (doubtless it is not a matter of strict logic that liquid spills out of overturned cups).

¹¹⁸ One can of course express probabilistic or “evidential” reasoning as strict logical inference if one wishes, so long as the “probably” operator remains present throughout (e.g.: If *x*, then probably *y*; *x*; therefore, probably *y*).

In mapping out this line of reasoning, I am not suggesting that this is the *only* way that the relevant inferences could proceed; I offer it as one way that they could proceed without the involvement of a cognitive attitude beyond belief and desire (and perception, if one grants there is a distinct perceptual attitude). That there are a number of inferences in the thought-process I will map out should not lead one to view this account as attributing especially complex mental processing. Most theorists do not go beyond formalizing a single inference involved in a pretense episode; here I try to map out a short episode in some detail. I will use P to indicate a perceptual content, B for belief, D for desire; if one is suspicious of a genuine distinction between perceptual and belief contents, a B can be replaced for each P without affecting the account.

A child's thought during a pretend tea-party:

P1: You say, "let's have a tea party," while holding an empty teapot and placing empty cups on a table.

B1: (*inferred from P1*): You want me to act, together with you, like we are having a tea party.¹¹⁹

B2: (*inferred B1 and from tone of voice, facial expressions, and other contextual clues*): We are not really going to have a tea party and you do not believe we are really going to have a tea party.¹²⁰

D1: (*caused by B1 and B2*): I act like we were having a tea party.

¹¹⁹ The child here attributes a desire to the experimenter; as Baron-Cohen *et al.* (1986) showed, even autistic children who have difficulties attributing beliefs to others are able to attribute desires. So, while two-year olds may not have an adult-like grasp of belief, it is not implausible to suppose that they readily attribute desires to others.

¹²⁰ I discuss below (Section Five) why the child's disbelief in the initial premise, and disbelief that the experimenter believes the premise, form necessary parts of the pretense.

B3: (*background beliefs or “scripts” concerning tea parties retrieved in order to help satisfy*

D1): People drink tea at tea parties; tea is poured from teapots into cups at tea parties.

P2: You make a pouring motion with the teapot over the two cups.

B4 (*inferred from B3 and B1*): You are acting like you are pouring tea out of the teapot and into the cups.

B5 (*from B4, due to D1*): I should act like you poured tea into both cups.

P3: You put down the bottle and say “watch this!”; you turn the green cup upside down and then put it back on the table, right side up.

B6: (*background beliefs*): When cups containing liquid are turned upside down, the liquid spills out of them. When full cups are not moved, they remain full. Tea is a liquid.

B7(*inferred from B5 and B6*): If you had poured tea into both cups and overturned the green one, the green one would now be empty.¹²¹

B8 (*inferred from P2, B5, and B6*): You are acting like you poured the tea out of the green cup.

P4: You say “show me which cup is empty and which is full.”

B9 (*inferred from P4, B1 and B2*): You want me to tell you which cup would be empty and which one full, if you had filled both cups and overturned the green one.

D2 (*caused by B8*): I tell you which cup would be empty and which one full, if you had filled both and then overturned the green one.

¹²¹ This belief might also be an indicative conditional (“If you pour tea into both cups and overturn the green one, the green one will be empty”). The counterfactual version seems more natural to me, however, as the indicative seems to imply that the child is *anticipating* that tea might really be poured into (and out of) cups, when there is no such anticipation or expectation. Rather, the child is trying to solve a kind of problem—to detect what the experimenter would be doing, if they were really having a tea party—to which consideration of counterfactuals seems more appropriate.

D2 then leads the child consult B7 and B8 in giving her answer: she points to the green cup to indicate that it is “empty”, and then to the other to indicate it is “full”.

This account assumes the child to have beliefs of the form “if x were the case, then y would be,” which there is good reason to think she has. These form parts of the “scripts”—the beliefs about how certain types of situations typically go—that Nichols and Stich appeal to as well. Moreover, Harris (2001) discusses a variety of empirical studies indicating that “young children have the competence for counterfactual thinking, spontaneously engage in such thinking, and deploy it in their causal judgments” (p. 252). In many instances, children are able to infer that if x had been the case, then y would be the case, and thus come to believe a counterfactual conditional.¹²² If they also have a desire to act as though x (where, e.g., x = a full cup was overturned), then this would lead naturally to a desire to act as though y (where y = the liquid spilt out), since acting as though y is just part of acting as though x .

Note that none of the beliefs I appeal to are “tagged” in any special way to indicate that they are mere “pretend” beliefs (Nichols and Stich rightly suggest that some other prominent proposals (e.g., Leslie, 1987; Perner, 1991, p. 53-67) that do “tag” certain beliefs in this way essentially reduce to their own view).¹²³ Nor are these beliefs conceptually onerous—they do not involve the concept PRETEND, or the intentional operator, ‘it is pretense that.’ This circumvents Currie and Ravenscroft’s (2002, p. 13-15) worry that the

¹²² For instance, Harris cites a study where 3-and 4-year-olds watched a Teddy bear paint the floor with a brush and were then asked both how the floor would look if Teddy had not painted it, and how it would look if he had painted it with his fingers instead of a brush. The children “were quite accurate in their replies, correctly anticipating what would have happened given the two types of counterfactual antecedent” (2001, p. 251).

¹²³ Perner (1991, p. 53-57) develops a theory of pretense that appeals to mental “models” (a “Reality Model, and an “As-if Model”) that serve to quarantine potentially contradictory representations (the representations guiding pretense reside in the “As-if” model). According to Perner “the representation controlling play action must be contained in a *different model* than information about the real world controlling serious action; otherwise, the child would be confused about what is real and what is pretend” (p. 55). If I am right that one can pretend that p without entertaining the proposition that p , Perner’s theory is overly complex for the same reason that N&S’s is.

use of such an operator in pretense reasoning would make pretending too cognitively demanding for children. Thus, there is no reason to doubt that two-year olds can engage in a wide variety of such pretenses.

So it does not appear that a special process need go on inside a PWB in order for a child to pretend, and to give correct reports about what is going on in the pretense. Keep in mind, when the cups are initially “filled” during the pretense, the child does not need to infer (or believe) *that the cups are full*; rather, she needs to infer *that the cups would be full, if liquid had been poured into them*. This latter belief poses no threat to any beliefs she has outside of the pretense.

This is obviously not how Nichols and Stich envision things. They suggest that, in order to form the beliefs relevant to guiding pretense behavior, one must access the description “unfolding” in the PWB; it is precisely by accessing these descriptions that one comes to have beliefs of the form, “If it were the case that p , then it would be the case that $q_1 \& q_2 \dots \& q_n$ ” (p. 128). Of course, there is nothing *in* the PWB other than the initial premise that begins the pretense-episode together with one’s standing beliefs (*minus* those that conflict with the premise). So it is hard to see what can be gotten *out* of it that cannot be gotten out of the Belief Box, so long as the initial premise can somehow cause appropriate processing to go on in the Belief Box without also causing inferential “chaos” or “representational “abuse”.

On my account, the initial pretense premise that- p causes processing in the Belief Box through the pretender’s desire to act as she would if p (again, this is a desire that N&S also grant pretenders must have). Having such a desire will cause her to search her memory for relevant beliefs or “scripts” about what would be the case if p . If one has no beliefs about

what would be likely if p , then the pretense won't get very far. But this is also true on N&S's account. If, however, one has some relevant beliefs—e.g., that tea is a liquid, that it is held in pots, that tea parties are cordial affairs, that tea spills out of overturned cups—then accessing these beliefs will give the pretender the resources for inferences relevant to acting in the way one would if p .

The assumption apparently driving “single code” accounts—N&S's, and many others—is that, in order to find out what would be the case if p , one needs to represent (in the PWB) that p is *the case* (which is just to represent that p), and then see what one comes to infer, as opposed to simply accessing whatever beliefs one has about what typically happens in situations where p . Harris (2001) explicitly voices this common assumption, when he notes that counterfactual reasoning guiding pretense depends “on the ability to temporarily entertain a representation that is non-veridical, and known to be so” (p. 252). But representing that p —be it in the PWB, or anywhere else—will result in relevant inferences being made only if one already has commitments about what is generally true of situations where p (or if one has beliefs about how situations go when q , together with beliefs that allow one to infer that p is a q -kind of situation). And if one already *has* those commitments, there is no need to represent that p in order to deploy them. One simply needs a desire to retrieve any relevant beliefs concerning what would be the case if p . Doing this does not require entertaining any non-veridical representations.

Another way to see the issue is from the perspective of simple hypothetical reasoning. According to N&S, it is the evolutionary function of the PWB to enable hypothetical reasoning (2003, p. 58). If I want to know what would happen if p while believing that not- p , then there needs to be a special “off-line” place in the mind where I can safely put the

proposition that p while cranking out whatever consequences rationally follow from it, given my other beliefs (Currie and Ravenscroft (2002, Ch. 2) espouse a similar view). But if reasoning about what would be the case if p does not require entertaining the proposition that p , then why think that the “offline” part is necessary?

Suppose I want to hypothetically reason about what will happen if the Cubs win the World Series this year (call the proposition that the Cubs win the World Series this year ‘ c ’). The desire that *I know what will happen if c* will be enough to cause me to access whatever general beliefs I have about teams that win the World Series. A few come to mind: I believe that the team jumps for joy (‘ j ’), that their fans shed tears of elation (‘ e ’), that they take part in a ticker-tape parade in their home city (‘ t ’), that shirts are printed up (‘ s ’), and so on. Having brought these generalizations to mind, and believing the Cubs to be the sort of thing that falls under those generalizations (i.e., a baseball team), I then infer that if c then j and e and t and s , etc., and form the corresponding beliefs. There is no need during all of this to put c itself in either the belief or desire “boxes”—or any “box” at all (hence, no need to “quarantine” c).

This makes the question of quarantining irrelevant; there is no need to prevent inferential chaos if contradictory propositions are not entertained in the first place. Nor is there any peculiarity in a person’s ability to imagine or pretend that p while not believing that p —or indeed while believing that p (though, for reasons given in Section Five, I deny that one can *always* pretend that p while believing that p). *For the activity of imagining that p consists merely in retrieving one’s beliefs in conditionals and generalizations relevant to the proposition that p , and using them to form additional beliefs about what would likely happen if p , all of which may (or may not) guide a sequence of pretend behavior (when they do guide*

such a sequence of behavior, the person moves from merely imagining that p to pretending that p). There is no reason to think that one's ability to do any of this would be hampered or confused by a concomitant belief that not- p , or that p .

As for the pretender's ability to distinguish what happened in the pretense from what actually happened, the main difficulty vanishes once one gives up the idea that the cognitive system must sort through contradictory propositions, or through multiple copies of the same proposition, in distinguishing the actual from the pretended.

4.3. *The Ramsey analysis*

For some, the contrary doctrine seems rooted in a questionable interpretation of the "Ramsey test" for determining whether a conditional should be accepted. Currie and Ravenscroft suggest that there is "nothing new" in their suggestion that imaginings are belief-like (though are not strictly speaking beliefs) because "philosophers interested in belief dynamics assume that imagining is belief like...when they offer the Ramsey test as a way of deciding whether you should accept a conditional" (2002, p. 12). They gloss the Ramsey test as:

Given an overall state of belief G , you should accept the conditional 'If P then Q ' if you should accept Q in the overall state $G \& P$, where $G \& P$ is the belief state that results from revising G in the light of P ¹²⁴ (p.12).

Since we cannot simply add P to our store of beliefs (because we cannot believe arbitrary propositions at will), determining whether we should accept 'If P then Q ' can instead be accomplished, they suggest, by adding P "in imagination...since imagination preserves the

¹²⁴ It should be added that the revision must be minimal or conservative.

inferential patterns of belief.” If Q then “emerges as reasonable” one knows (via a putative implementation of the Ramsey test) that ‘If P then Q ’ should be believed.

To call Ramsey’s proposal a “test”, however, is misleading. What Ramsey is offering (if he is offering it¹²⁵) is a *reductive analysis* of the conditions under which we typically accept conditionals—one which reduces acceptable conditions for assenting to a conditional to acceptable conditions for accepting various individual propositions. This sheds some philosophical light on the nature of conditionals, and provides a yardstick for testing the plausibility of other philosophical accounts of conditionals.¹²⁶ He is not, however, providing a “how to” guide for determining whether one should accept particular conditionals—that is Currie and Ravenscroft’s invention. One might wish to *argue* that Ramsey is providing such a test. But, as long as hypothetical reasoning can proceed as I have described above, the Ramsey “test” can be accepted at face value without any need to conclude that a distinctive propositional attitude is engaged when deciding whether to accept a conditional. It is simply through considering my standing beliefs about what generally happens when teams win the World Series that I come to accept the conditional, ‘If the Cubs win the World Series, they will appear in a parade.’ What justifies my accepting the conditional is whatever justifies my beliefs that teams that win the World Series appear in parades, and that the Cubs are a baseball team. Again, there is no reason to think that combining such beliefs in order to make this inference cannot take place solely within the Belief Box—for I do not need to form any arbitrary beliefs, nor need I entertain a proposition that conflicts with any of my beliefs.

¹²⁵ Currie and Ravenscroft note that the Ramsey test they describe “may not be his own”; whether it really is Ramsey’s idea is not important for present purposes—C&R’s understanding of it is common enough to warrant discussion.

¹²⁶ Cf. Jonathan Bennett’s (2003, pp. 28-33) discussion of the Ramsey test and its relation to theories of conditionals. Bennett argues that the Ramsey test showed the truth-functional “horseshoe” analysis of the truth conditions for conditionals (advocated by Grice (1989), among others) to incorrectly characterize the indicative conditionals of natural language.

4.4. *Explaining inference – mirroring and diverging*

Another important question under the rubric of ‘inference’ was what accounts for the “inferential orderliness” of the various inferences made during pretend play—why is it that what we infer of pretend scenarios so closely mirrors what we would infer if we believed the pretend premises? Nichols and Stich account for this by positing that the same inference mechanism works on representations in the PWB as in the Belief Box—an identity of mechanism enabled by the representations in each box being “in the same code.” On my view, the answer is simpler: the beliefs concerning conditionals and generalizations that guide inferences and behavior in a pretense are generally *the very ones* that guide inferences and behavior in real life. For example, if one is told, “pretend that Bob was in New York on Monday, and London on Tuesday,” one will typically infer, as part of the pretense, that Bob got to London via airplane, just as one would infer that he’d gone by airplane if one were simply told that Bob was in New York on Monday and London on Tuesday. This “mirroring” is due to the fact that we tend to fill out pretend and actual scenarios by appeal to the same beliefs about how things normally go; in this case, the relevant belief is that people who travel that far that fast usually do so by airplane.

Why do we do it this way? Why does imagining that *p* feed off beliefs concerning what would be the case if *p*? Here I think we have a pseudo-question; imagining that *p*, on my view at least, *just is* bringing to mind beliefs concerning what the world would likely be like if *p*, usually by drawing inferences from other beliefs one deems relevant. The question of why the inferences drawn “in imagination” mirror those that would be drawn from “isomorphic” beliefs is puzzling only if you begin with the view that imagining is not itself a

way of believing, and that the propositions involved in imagining are quarantined in their own special box. Only then will it seem attractive to attribute the (non-existent) mirroring to a mechanism that treats the representations in both boxes in roughly the same way.

Of course, imagining that p and pretending that p tolerate a good deal of divergence from our beliefs about what would likely happen if p , and from what our actual behavior would be if p . It might be thought that the *creativity* of imagination—our ability to imagine things we would never anticipate actually occurring—poses a problem for an account which assimilates imagining to retrieving and inferring various beliefs about what would likely be the case in certain scenarios. However, creativity poses no particular problem. The creative freedom of imagination derives from our ability to insert a new premise into our imaginative projects whenever we wish, and to draw out further inferences from it, where “inserting a new premise” n amounts to having a desire to determine what would happen if n .

For example, if I start out imagining that I am walking down the street, this will consist in entertaining various beliefs about what it would be like for me to be walking down the street. If I imagine that the normal things that happen when a person walks down a street *don't* happen, then this consists in forming beliefs about what would happen if the normal things that happen when a person walks down the street don't happen. If I imagine myself leaping over a building in a single bound, this will consist in forming beliefs about what would likely happen were I to leap over a building in a single bound (e.g., what the building would look like as I flew over it, what onlookers would exclaim, etc.). So, to the extent that I can form some beliefs about what the world would likely be like if p , I can imagine that p .

Again, we should not be distracted by the fact that one can imagine abnormalities; for instance, one can imagine that a bowling ball bounces high into the air after being dropped.

This is not forming beliefs about what the world would likely be like if a bowling ball were dropped; but it *is*, I claim, forming beliefs about what the world would likely be like if a bowling ball bounced high into the air. We need posit no special process or cognitive mechanism by which new and unusual premises are freely selected and used to extend an imaginative project (no “Script Elaborator”) –the process is simply driven by our desires and interests with regard to the pretense or imagining (do we want to know what will happen if it rains at the parade? Do we want to know what would happen if the sky turns red?, etc.). The freedom of imagination is just a special case of the more general freedom we have to reason and think about topics of our own choosing.

5. Currie and Ravenscroft’s “belief-like” imaginings

I turn now to another prominent discussion of imagination—that of Currie and Ravenscroft (2002). They share with Nichols and Stich many of the same views of imagination and its relation to belief. In “single code” fashion, they argue that “belief-like imaginings” and beliefs can have the same content while differing in attitude, and that there are important similarities between belief-like imaginings and beliefs, consisting mainly in the belief-like inferential patterns of (propositional) imaginings (2002, Ch. 1).

Their theory is rather more impressionistic than Nichols and Stich’s; in lieu of a PWB and Belief Box, they posit “a vaguely bounded irregular solid” containing one’s various belief states, and another “irregular solid” including belief-like imagining states corresponding to each of those belief states (pg. 16). And, in lieu of a single inference mechanism that works on both boxes, there is a single “decision-making system,” that processes both kinds of states, but into and out of which imaginative states have a “special

pathway” and “sideways exit” (p. 68). The most substantive difference in the views is that C&R posit “desire-like imaginings” (which bear to desires roughly the same relation that belief-like imaginings bear to beliefs), while N&S do not.

Like N&S, C&R are impressed by the similarities of imagining and believing, yet they are also keen to show that imagining cannot simply be a kind of believing. Their arguments have a more conceptual flavor to them than Nichols and Stich’s; considering them will, I think, help clarify my own position.

One way C&R think belief differs from imagination is that “imaginings are doings and believings are not” (p. 15). The intuition here is that we actively control what we imagine, whereas we cannot control what we believe; our beliefs simply appear on our doorstep, for good or ill.

Yet, one might grant that we cannot form whatever beliefs we wish, while nevertheless holding that certain ways of entertaining beliefs count as “doings”. Remembering is a clear case of a mental action—a mental “doing”—that simply involves belief. Drawing inferences through comparing one’s standing beliefs is similarly a “doing” that may only involve beliefs. On my view, when we imagine things, we are (“actively”) remembering, reasoning, and inferring; the freedom of imagination comes in our ability to choose the subject matter of this thinking—to take whatever wild premise (or premises) we wish and draw out the consequences, and to freely add new premises as we go.¹²⁷ In short, there are many ways of viewing believing as a “doing”, while denying our ability to arbitrarily form beliefs; imagining can constitute one of these “active” ways of believing—or several of them simultaneously.

¹²⁷ See my more extended discussion of this above in Section 4.

Currie and Ravenscroft's main argument for denying that imaginings are beliefs, however, is a “normative” difference they see between the two:

Belief, however weakly characterized, is normative in that an agent who has contradictory beliefs (in any sense of ‘belief’) is in a less than ideal epistemic situation. It is no defect in an agent’s epistemic condition that she imagines things contrary to what she believes (17). This is essentially just another way of putting the claim that one can unproblematically imagine that *p* and believe that not-*p*. I bring it up again because it is explicitly offered here as a reason—really, *the* reason—for denying that imagining can be assimilated to believing. What C&R seem not to appreciate is that their argument simply *assumes* that imagining that *p* involves taking a certain attitude toward the proposition that *p*; they question only whether this attitude is belief, not whether the proposition that *p* is entertained. It is only once the latter assumption is in play that the normative considerations they cite have any force.

Of course, I reject the assumption that imagining that *p* requires one to entertain the proposition that *p*, and so propose an end run around this popular line of argument. If single code theorists really want to *argue* that imagining is not believing, they need to provide some reason for thinking that when someone has imagined that *x* he must entertain the proposition that *x*. I have already argued that the human facility for pretense does not force the conclusion that imagining that *p* requires entertaining the proposition that *p*, or the positing of a distinctive imaginative propositional attitude.

There are, however, two other phenomena arising out the empirical literature that single code theorists often count in favor of their approach: the co-presence of pretense and mindreading deficits in autism, and the effect of imagining and pretending on affective and emotional systems. I turn to these now.

6. The link to mindreading

The empirical and conceptual issues surrounding autism are vast and fascinating. As this chapter is principally an assessment of whether pretense requires a distinct cognitive attitude, I will only have space to indicate why neither N&S nor C&R's theories are bolstered by a potential extension to mindreading. Both Nichols and Stich (2003) and Currie and Ravenscroft (2002) argue that their approaches to pretense—insofar as they invoke a distinct cognitive attitude—are supported by their ability to explain mindreading as well. Just as pretense, on their views, requires an ability to entertain and draw “belief-like” inferences from propositions one does not believe, so too does mindreading require the ability to take another’s potentially different beliefs into account, drawing out implications that will help predict that person’s behavior. If both abilities are sometimes simultaneously impaired—as appears to be the case in autism—this would suggest that a single system underlies both.

People with autism have difficulties in three main areas: pretense and imagination, social competence, and communication. These make up ‘Wing’s triad’ of autistic deficits (Wing & Gould, 1979). A number of theories have been developed to explain why these three disabilities appear together; one of the most influential derives from the work of Simon Baron-Cohen (1989, 1995), who traces the deficits to a general inability to conceptualize and understand the mental states of other (i.e., to “mindread”). This hypothesis is grounded in the striking fact that autistic children consistently fail “false belief” tasks that are easily passed by unaffected children of the same verbal age and linguistic competence. In a standard false belief task, the subject watches an experimenter (or a puppet) place an object in a box or container. The experimenter then leaves the room and an accomplice takes the

object out of the container and places it in another one nearby. The subject is then asked where the experimenter will look for the object when she returns. Subjects who indicate that the experimenter will look in the container where the object was first placed are taken to understand that the experimenter has a false belief about where the object now is. This, in turn, is taken to show that the subject has an ability to think about the mental life of another person.

As remarked, children with autism usually fail false belief tasks that are passed by non-autistic children of the same mental/verbal age. They also, in keeping with the main symptoms of autism, rarely spontaneously engage in pretend play. Nichols and Stich think that these deficits “might best be explained by the hypothesis that something is damaged in the PWB system,” since they take themselves to have already established that the PWB system “is implicated in sophisticated mindreading and in pretense” (2003, p. 129). Currie and Ravenscroft analyze autism similarly as a “disorder of imagination,” where the imagination is “a device which assists us in understanding and solving problems across a range of different domains,” and is “heavily employed in helping us understand and respond to the thoughts and feelings of other people.” They find it “an advantage of our approach” that there are other domains (pretense and planning) where autistic people show deficits and where it is “antecedently plausible that...imagination is heavily used” (2002, p. 139). (Note that both N&S’s and C&R’s accounts of autistic deficits rely upon the claim that the deficits involve mechanisms we have antecedent reason to believe exist—the antecedent (and disputed) reason being that they are required for pretense).

The main problem with both N&S and C&R’s explanations when extended to autistic deficits is that they are much too broad. On both accounts, not only does the PWB (or

“belief-like imaginings” for C&R) underlie our ability to engage in pretense and mindreading, it also allows for hypothetical and counterfactual reasoning. On N&S’s theory, the “‘proper function’ of the pretense mechanisms” is not simply mindreading and pretense, but “more broadly for hypothetical and counterfactual reasoning,” (2003, p. 47); and C&R hold that supposing, or assuming for the sake of argument, involve belief-like imaginings (2002, p. 33-38). The problem is that hypothetical and counterfactual reasoning (and “supposing”) *per se* are not impaired in children with autism.

Indeed, Scott et al. (1999) found that autistic children outperformed normal children of matched verbal age on counterfactual reasoning tasks. Interestingly, the performance of the autistic children declined only when they were prompted to form visual images while considering their answers to the questions.¹²⁸ Peterson and Bowler (2000) showed that autistic children have a facility with a kind of hypothetical reasoning they call “subtractive reasoning”. Normal false-belief tasks can be rephrased using “subtractive” prompts such as, “If the marble had not been moved, where would it be now?” in lieu of “Where does Sally [falsely] believe the marble to be?” When the tasks were rephrased in this manner, autistic children were able to provide correct answers at levels comparable to those of their non-autistic peers. Further, children of all kinds they studied—normal, autistic, and learning-disabled—who could not answer the subtractive questions also could not answer the false-belief questions. This led Peterson and Bowler to conclude that subtractive hypothetical

¹²⁸ These results have been criticized (Leevers & Harris, 2000) as being due to a bias of autistic children to answer “Yes” to questions (the correct answer to each of the questions was in fact Yes). Yet, as Scott and colleagues point out, this fails to explain why their answers became considerably less accurate (involving many answers of No) once they were encouraged to form images while answering the questions (the correct answers were still all Yes).

reasoning, preserved in autism, is a necessary but not sufficient factor in accurately answering false-belief questions.

In an earlier study, Peterson and Bowler found that autistic children respond appropriately to explicit counterfactual questions, such as “If Mummy hadn’t made the cake, where would the chocolate be?” (Peterson & Bowler, 1996); and Hadwin and Bruins (1998) found that autistic children could formulate counterfactual antecedents and consequences for various episodes. For instance, one child suggested that by wearing boots a story character could have prevented getting her socks muddy. Moreover, we should remember that the mindreading deficits of autistic children are of special interest in large part because their abilities to make hypothetical predictions using other commonsense theories—in particular, “folk physics”—is generally intact (Baron-Cohen, Leslie, & Frith, 1986).¹²⁹ N&S’s theory does not predict that the PWB has more difficulty digesting folk-psychological generalizations than folk-physical ones. Thus, far from supporting N&S and C&R’s theories, the hypothetical and counterfactual reasoning abilities of autistic children present a serious challenge.

Of course, N&S can respond that it is not a problem with the *entire* PWB system that leads to the difficulties in autism. Rather, there may be problems with only those aspects of it that enable mindreading and pretense (and C&R may reply that there are only problems entertaining *certain kinds* of belief-like imaginings). But this is to remove any independent explanatory power of the PWB to explain autistic deficits. For then it is nothing about the PWB *per se* that explains the triad of difficulties; rather, it is some special defect(s) of this

¹²⁹ Some high-performing autistics (e.g. mathematical savants) even show a pronounced superiority in counterfactual reasoning tasks over the general population. See Baron-Cohen *et al.* (1999) for a discussion of several such cases.

system that explains the problem. Without antecedent reason to believe in the PWB, one can just as (un)informatively posit defects in the ability to understand other people's beliefs as the "explanation" of the deficits.

7. *Disbelief: a further condition on pretense*

While I have no competing theory of autism to offer, I do have a suggestion for why pretense and mindreading are simultaneously deficient: both require at least a minimal understanding of belief (what I have no answer to is why people with autism have difficulty understanding belief in the first place). That mindreading requires understanding belief has already been made clear. In my earlier discussion of pretense, I more or less went along with N&S's suggestion that "to pretend that p is...to behave in a way that is similar to the way one would (or might) behave if p were the case," and that "a person who wants to pretend that p wants to behave more or less as he would if p were the case" (p. 128). Yet now I will argue that these characterizations leave out a crucial aspect of pretense, which is that any pretense episode must involve us pretending something is case that we believe *not* to be the case (even if, within the pretense, many things are pretended that are also believed). I will then use that conclusion as part of a further argument to establish that pretense requires a minimal understanding of belief.

Suppose you are seated in a chair, minding your own business, when suddenly someone approaches and says: "Pretend you are sitting in a chair!" What should you do? While you sit there wondering how to comply (do you assume a comical posture?...no), the person says, "Now stop pretending!" You look at him. He says, "Now resume pretending!...Now stop again!"

These requests strike my ear as absurd.¹³⁰ The absurdity lies in the fact that you cannot possibly comply with the request—not so long as you believe you *are* sitting in a chair. For the pretense to get off the ground, it must extend into the realm of what you disbelieve; you must pretend that you are the King of Spain sitting in a chair (or, if you believe you are the King of Spain, something else).

We can approach the issue from another direction. Recall that, according to N&S, it is a desire to act as one would if *p* that motivates pretense behavior. But, usually, when we believe that *p*, we *also* desire to act as we would if *p*. When I believe that I am driving a car, I normally wish to act as if I am driving a car.¹³¹ So, acting in accordance with such a desire certainly isn't sufficient for pretense behavior (nor is acting as one would if *p* when not-*p*, since this fails to distinguish pretense from the behavior of someone who is genuinely mistaken as to whether *p*). Something else is needed in addition—namely, the pretender must be pretending that something is the case that she believes not to be the case.

To put the claim formally, *an act of pretending that p can begin only when the aspiring pretender wants to act as she would if p, while believing that not-q, for some q that partly governs the pretense (where q may or may not = p)*. To be sure, once this *q* is in place, one can go on to imagine, as part of the scenario governed (partly) by *q*, that various things are the case that one also believes to be the case (or that one neither believes nor

¹³⁰ The absurdity is removed if one is asked to pretend certain things *in addition* to “I am sitting in a chair” (e.g., “I am sitting in a chair and I am the king of Spain”). Here I am focusing on the request, “Pretend you are sitting in a chair—and pretend nothing else in addition.” Note that this is easily done by someone *not* sitting in a chair—and, indeed, by someone who *is* sitting in a chair but believes himself not to be.

¹³¹ Unless, of course, “act as if” is supposed to *mean* “pretend”—but it is not, on N&S’s view, since they stress that pretenders need not have the concept PRETEND.

disbelieves). Thus, as part of the tea party pretense, it is possible for the child to both believe and pretend that the green cup is empty. But theorists, such as N&S and Leslie (1994), are rash in concluding from such phenomena that pretending is in no way constrained by what we disbelieve. To get off the ground, an episode of pretense must in some way dip into the realm of what is disbelieved.¹³²

A bit more can be said about *q*; *q* “partly governs” the pretense in virtue of being a premise from which inferences guiding pretend behavior are drawn. In my terms, this means that as part of a pretense the pretender tokens a desire to act as he would if *q* (while disbelieving *q*), and this desire leads to the tokening of beliefs concerning what would happen if *q* that in turn drive some of the pretend behavior. The actions that together constitute the pretense *count* as pretense precisely because *q* is both disbelieved and used as a premise to guide the actions.

The connection of pretense to disbelief is important for three reasons. First, it reinforces the idea that the mental components of pretense can indeed be reduced to the tokening of ordinary beliefs (where some are disbeliefs) and desires. For if pretense were not at all constrained by what is disbelieved, yet still involved only belief, it would be hard to see how pretending that *p* could be distinguished (at a cognitive level) from mistakenly acting as though *p*. This might then motivate one to think that a belief-independent cognitive attitude *must* be involved in imagination.

¹³² I do not know of any way to test the claim that pretense requires pretending something that is disbelieved than to consider specific cases. Here is one that might appear to challenge the claim: I buy a Lotto ticket and, not knowing whether I have won, pretend that I have won. Here I neither believe nor disbelieve that I have won. But we need not look far for the proposition that is pretended and which is disbelieved—in pretending I have won, I am (as part of the same pretense) also pretending that I have found out from some reliable source that I have won, while believing that I have not.

Secondly, and relatedly, the requirement of disbelief in q reveals an independent problem for N&S's account. For if, as they argue, the imaginings driving pretense are quarantined from belief, and if pretense is simply behavior driven by such representations (which are themselves "copied" from the Belief-Box), there is no reason to expect any constraint on what we can pretend by what we (dis)believe. We should be able to put all and only believed propositions into the PWB and use them to drive a pretense; that is, we should be able to pretend the world is exactly as we believe it to be. But, like the sitting-in-a-chair example above, that seems absurd. Appreciating this absurdity should lead us to consider anew whether the cognitive systems underlying imagination and belief really are as distinct as N&S imagine them to be.

Thirdly, the requirement of disbelief in q also reveals why pretense requires an understanding of belief and, hence, why those with a poor understanding of belief (such as autistic children) will less easily engage in pretense. To see why, first note that pretend behavior in line with this new definition does not in itself require one to understand—or to have the concept of—belief; it merely requires that one *have* a belief in conflict with a premise of the pretense. The piece of the puzzle I now add is that, as Leslie (1994, p. 221-222) emphasizes, the abilities to engage in group and solitary pretense emerge at the same time; why this should be so is an explanandum for theories of pretense. The reason for the linkage, I suggest, is that children learn how to pretend on their own by watching others engage in pretense; and understanding that another person is pretending requires understanding that that person does not believe that p , for some p that governs the pretense—and *this* requires at least a minimal understanding of belief.

To see this, consider the famous banana-telephone example (Leslie, 1994; Nichols and Stich, 2000), where child and mother take turns pretending that a banana is a telephone. As both Leslie and N&S note, before a child understands pretense, it might seem very odd to see Mommy using the banana as a telephone. Given that children as young as two find nothing odd about such behavior, and often join right in, what is it that they understand that removes the *prima facie* oddity of Mommy treating a banana like a telephone? Nichols and Stich suggest it is that the children have a “behavioral” understanding of the pretense, consisting in their understanding “that Momma is behaving in a way that would be appropriate if the banana were a telephone” (2000, p. 139). They stress that this behavioral understanding of pretense is possible “without understanding that the other person has any mental states at all” (*ibid.*).

But this sort of “behavioral” understanding really isn’t sufficient. The very thing that is potentially odd about Mommy’s behavior is *that she is behaving in a way that would be appropriate if the banana were a telephone!* Understanding that she is doing this obviously cannot be what constitutes understanding that she is pretending. Another necessary condition must be met, namely that the child understand (i.e. believe) that Mommy does not believe that the banana is a telephone. Until the child believes *that*, Mommy will seem in sore need of a nap.

But, comes the objection, normal children do not pass false-belief tasks until the age of four, whereas they engage in pretense at two. If pretense requires an understanding of belief, shouldn’t two-year-olds be able to pass false belief tasks as well? The answer is that false belief tasks require a much more sophisticated understanding of belief than pretense does, since it requires children to understand that someone else can have *different* beliefs—

beliefs the child takes to be false—than the child herself has. Pretense only requires the child to recognize that, just as she does not think the banana is a telephone, so too does Mommy not really believe the banana to be a telephone. So, pretense requires an ability to *recognize a shared belief*, not to other-ascribe a contrasting belief.¹³³ Judgment can be reserved over whether this constitutes a full understanding of belief; certainly it lies along the road that leads to a full “folk” understanding of belief. This is a road that, for whatever reason, people with autism have difficulty travelling.

8. *Affect and Imagination*

A last consideration often thought to weigh in favor of single-code accounts arises from the affective and emotional responses of imaginers. When watching a film or reading a novel, we often have affective responses similar to the ones we would have if we believed that the fictional situations were real. This constitutes a well-known puzzle in aesthetics: why do we have strong feelings about things we do not believe to have happened? Similar responses occur outside of our engagement with artistic works as well. Nichols (2006) and Currie and Ravenscroft (2002) cite a range of empirical studies (many of which are reviewed in Lang (1984) and Harris (2000)) where subjects who are asked to imagine certain (often anxiety-producing) situations show at least mild versions of the affective responses appropriate to the situation being real.

¹³³ Leslie (1987), who also thinks that pretense requires a primitive mindreading capacity, accounts for the two year lag between pretense behavior and the passing of false-belief tasks by appeal to the greater overall complexity of the computations needed to pass false-belief tasks (p. 422-423). See also his (1994, p. 228-237) for a discussion of empirical studies in support of this idea.

A key claim of N&S's theory—and of single code theories in general—is that certain cognitive mechanisms are “source-indifferent”, processing input from multiple sources (e.g. the Belief Box and the PWB) in “much the same way” (Nichols, 2006b; Gendler, 2006; Weinberg & Meskin, 2006a). These mechanisms can be thought of as abstracting away from attitude, operating only on content (Currie & Ravenscroft, 2002). One such mechanism, already discussed, is thought to be the “inference mechanism” responsible for drawing inferences from beliefs in the Belief Box (and, allegedly, from imaginings in the PWB). A potential explanation of our affective responses to artworks, and to imagined scenarios in general, is that the mechanisms giving rise to affective and emotional responses are similarly “source indifferent”, or “attitude-neutral”, accepting and processing input from the PWB and Belief Box in much the same way (while other mechanisms, such as those controlling action, do not take input from the PWB¹³⁴). This purports to explain why imagining that *x* can give rise to affective responses similar to those that arise from believing that *x*, without giving rise to similar behavior.¹³⁵

A rather large problem with this suggestion is that our affective responses to what we imagine are hardly ever of the same order as they would be were we to believe what we are imagining, and very often the disparity is extreme. The anxiety and despair I feel when imagining that the Empire State building has been hit by commercial jets and has crumbled

¹³⁴ Though see Gender (2006) for cases where it does—she calls such violations of quarantining “contagion”. Incidentally, there is no need to explain “contagion” if there is no quarantining to begin with. There are many reasons that forming beliefs about conditionals and counterfactuals (what I take imagining to be) would occasionally guide or influence behavior.

¹³⁵ See Goldman (2006, p. 45-46) for an important criticism of this proposal. His argument, in brief, is that the mere fact that affective response mechanisms *can* receive input from N&S's PWB (in virtue of such representation being in the same code as Belief Box representations) does not make it likely that they *would*. For a theory's posits to be explanatory, “the explaining facts (the *explanans*) should imply, or at least make probable, the fact to be explained (the *explanandum*)” (p. 45-46). As he points out, desires are “in the same code” as beliefs, but this does not imply or make it probable that they would feed into the same “attitude neutral” affective response mechanisms. Why, then, should we *expect* the representations in the PWB to do so?

to the ground, killing thousands, is nothing at all like the affective response I actually had when I came to believe that the World Trade Centers were destroyed; it is closer to—though not yet as intense as—the sinking feeling I had a moment ago when I realized that I have to do a load of laundry this afternoon. Nor do I feel any elation or any sense of fulfillment when imagining myself accomplishing important personal and professional goals, no matter how detailed the fantasy; I can only hope that this does not in itself show that such elation and fulfillment will never come.¹³⁶

But let not my own fallible introspection purport to settle this matter. Looking at the one specific study discussed by Nichols (2004; 2006), Lang *et al.* (1983) found that snake-phobic individuals exhibited physiological fear responses (elevated heart rate and increased skin conductivity) when imagining an encounter with a snake similar to those they exhibited when actually encountering a snake. Moreover, snake phobic individuals reacted more strongly to imagining an encounter with a snake than did non-snake phobic individuals, suggesting that such responses mirror one's actual responses in a fairly fine-grained way (Nichols is particularly impressed by this aspect of the results, as it is consistent with the idea that one's own beliefs are copied into the PWB). Yet there is another aspect of this study that Nichols fails to mention: these “mirroring” responses were only present after an extensive “imagery pretraining program” where subjects were coached in relaxation and visualization techniques relevant to the task. In another experiment that was part of the same study, but which lacked this “response training” regimen, Lang *et al.* report that “neither

¹³⁶ Nichols and Stich could respond that such disparities are due to the fact that *only some* affective and emotional systems take inputs from both belief and emotional systems (Nichols (2004) contemplates this response). However, this move prevents the phenomenon of affective response from presenting an independent consideration in favor the single code hypothesis. In any event, Nichols (2006) does not adopt this strategy in his detailed response to this challenge.

group generated a significant physiological reaction to the fear content during imagery” (p. 276).¹³⁷

It is hard to see why such a training period would be necessary on the PWB account, since the mere act of supposing thus and such to be the case (without being coached to form detailed imagery) constitutes imagining for N&S, and there is no reason to think individuals would have an initial difficulty supposing that they are encountering a snake (recall again the link between the PWB and hypothetical reasoning). When considering any such study revealing affective and emotional responses to imaginary scenarios, we must keep in mind that the fact that such reactions *can* be elicited in certain instances (just as they *can be* elicited by compelling artworks) does not in itself support the single code hypothesis. For the single code hypothesis predicts that all (or at least *most*) cases of imagining (or supposing) that *p* will result in affective responses very similar to those attending the belief that *p*. Since there are many cases where no such mirroring occurs, the single code hypothesis is *challenged*—not supported—by the diversity of affective response.

Nichols (2006) has recently responded to this line of criticism. He appeals to the fact that the inferences we draw from a single premise, both in belief and in imagination, are guided by our actual interests and desires. We sometimes have different interests when we imagine something (e.g., to find out what a Utilitarian would say) than we would if we believed it, and this can lead to our “filling out” the imaginary scenario in different ways (esp., including fewer alarming details), resulting in different final input to the affective

¹³⁷ Another reason to treat this particular study cautiously is that it concerns individuals with phobias—i.e. *abnormal* fear responses. Such responses to merely imagined scenarios may constitute part of a psychological *disorder*, not the normal feeding of PWB contents into an affective mechanism.

mechanism that still—in accordance with single code theories—treats the PWB’s input much the same way that it treats input from the Belief Box. Here is Nichols:

The explanation for the asymmetries [in affective response] is *not* that the affective mechanism itself responds differently to imagining that *p* and believing that *p*. Rather, the asymmetries arise because the affective mechanism is sent quite different *input* depending on whether one imagines that *p* or believes that *p*...in the cases of absent affect, what we are imagining is extremely sparse. (2006, p. 472)

So, according to Nichols, we can imagine disastrous events without generating affective responses because, when we do so, we do not have a desire that pushes us to fill out the scenario in a way that would duplicate the kinds of inferences that would be drawn if we believed the disaster to have really happened. It’s the details that hurt, in short, and thankfully we leave most of these out when we imagine disturbing events.

This suggestion is not very plausible. While we may sometimes engage in the kind of “sparse” imaginings that Nichols contemplates, we are certainly not obliged to. As he stresses, “we have vast control over what gets *into* the imagination” (2006, p. 469). So, just as we can keep alarming details out of what we imagine, we can also, *if we so choose*, fill in the imaginary scenario with arbitrarily many ghastly details. Indeed, when I imagine that the Empire State Building is being destroyed in a 9/11 style attack, there is no shortage of ways I can fill in that scenario, since I unfortunately have at hand a full stock of beliefs about how such an event would play out. Nevertheless, the kinds of affective responses thereby generated are hardly comparable to those that would be generated by believing the event to have occurred.

This difficulty in accounting for affective responses (and lack thereof) to what we imagine might seem a mere annoyance for the single code hypothesis if one was otherwise strongly convinced of its merits. But if, as I have argued, neither pretense nor mindreading data favor the theory, it is hard to see how phenomenon of affective response is anything other than a challenge to the hypothesis.

Why, on my account, is there any link at all between imagination and affective response, if all we are doing is reasoning hypothetically? This is not the place for a full-fledged account, but I will point in the direction of what I suspect is the right answer. According to me, imagining that a snake is about to bite involves, among other things, tokening various beliefs about what would happen if a snake bit me. There is obviously no general problem understanding how beliefs can trigger affective response, but why would beliefs about mere possibilities and conditional connections trigger affective response? As avoiding snake bites is adaptive, there is an evolutionary advantage to my experiencing at least some negative affect when considering such a possibility. The presence of affect means that I do not need to carry out a logical deduction from premises concerning what I desire (to survive), together with the likely effect of snake bites on the achievement of those desires, to be motivated to avoid poisonous snakes; the anxiety triggered by consideration of the mere possibility of being bitten is enough to motivate avoidance behavior. Of course, it would seriously hamper my ability to lead a (re)productive life if I were paralyzed with fear at the consideration of dangerous (mere) possibilities—hence the more mitigated affective response.¹³⁸ Analogous points can be made in the case of the positive affect felt at the

¹³⁸ There are no doubt many cases where imagining things that would negatively impact one's reproductive "fitness" does not give rise to negative affect. However, as long as there are very many cases where imagining

consideration of conditionals representing positive possibilities. Such imaginings can play a motivating role by providing a taste of the full emotional payoff that would ensue from realizing the possibility—without providing so much comfort as to remove any desire to attain the goal.

Again, this is only the barest sketch of a proposal for dealing with affective responses to imaginings. It likely raises as many questions as it purports to answer. For one, it assumes an account of how emotions interact with and influence cognitive states like beliefs and desires—an account I will be unable to provide. Nevertheless, much of the mystery concerning affective responses to imagined scenarios is removed simply by recognizing that imagining principally involves belief, and not some other potentially disconnected cognitive attitude.

9. Sensory and Propositional Imagination United

I have argued that we need not posit a PWB, or a distinctive cognitive attitude, to account for pretend play in children and adults. Such activity can be better explained by appeal to ordinary beliefs and desires. Nor does the data from autism, or the phenomenon of affective response to the imagined, weigh in favor of Nichols and Stich's (or Currie and Ravenscroft's) single code theory, once a single attitude account of the kind I have argued for is available.

My view is that propositional imagining is *forming a belief about* what would be likely in various situations, through consideration of one's standing beliefs concerning relevant generalizations. Pretense consists in using such beliefs to guide one's behavior

harmful possibilities (e.g., falling off a cliff, being attacked by a mountain lion) does trigger such affective response, an appeal to the evolutionary advantageousness of such responses can succeed.

situations where one is acting as though something is the case that one believes not to be the case.

In closing, I want to discuss an important lingering issue: the relation between propositional and sensory imagination. While propositional imagination—the subject of this chapter—is commonly thought not to *require* mental imagery, it is plausibly still often *attended by* imagery; I can sit and imagine *that we are having a tea party*, and form lots of images while doing so. Then I will also be (non-propositionally) imagining *a tea party*. And while some may grant that the non-imagistic components of propositional imagination may simply be the kinds of beliefs I have suggested, there may be more resistance to the claim that imagery itself is a kind of belief. Could it be that the imagery involved in acts of propositional imagination is generated in something akin to a PWB—but for quasi-perceptual states?¹³⁹

The first three chapters of this dissertation already furnish my answer to this. I hold that sensory imagination itself consists of doxastic, assertoric states that (roughly) represent the way things looked or would look. Thus I do not think of imagery as occurring “off-line” any more than I think of the non-imagistic states involved propositional imagination occurring off-line in a PWB. The mental representations guiding pretense and constituting imagination are all on-line (poised to guide action), and none involve acts of willful misrepresentation—that is, of representing things to be as one believes them not to be.

I have already given the arguments for this view of sensory imagination in detail; I won’t repeat them here. However, as a way of clarifying my overall view of propositional imagination, I’ll now sketch out the mental life—as I imagine it—of someone engaged in an

¹³⁹ See Currie and Ravenscroft (2002, ch. 4) for a view of this kind.

act of propositional imagination that is attended by mental imagery (and hence is partly constituted by acts of non-propositional imagination). It should be emphasized that, on my view, propositionally imagining that *X* is not a matter of tokening a single imaginative mental state with the content that *X*, but is rather a diachronic mental activity that usually involves the tokening of a variety of mental states, some of which may be states with the kinds of contents I argued in Chapter 2 are distinctive of sensory imagination. Just as imagining that *X* may involve forming a belief that if *r* then *q*, it may also involve a state of the form: IMG PM (there could be an *X* with thus and such visible features AND *Y* would look roughly this way).

So then, suppose I want to imagine (propositionally, but with attendant imagery) that *60 Minutes* correspondent Andy Rooney is President of the United States. For simplicity, I will break this imaginative project into three imaginative acts:

- 1: Imagining Andy Rooney hunched over his desk in the Oval Office (note that this imagining is non-propositional, indicating the presence of imagery).
- 2: Imagining that he gains a facility with tele-prompters.
- 3: Imagining that bills in obvious conflict with common sense are vetoed.

We can break each of these three “imaginative acts” further into assertoric states of various kinds; while 2 and 3 could conceivably characterize imaginative projects lacking any imagery, I will assume that they here are accompanied by imagery, just as 1 is. In connection with 2, I might imagine Andy Rooney glancing at a clear rectangle. In connection with 3 I imagine a red stamp marked VETO slamming down. Behold, then, the inner life of an imaginer:

Imagining that Andy Rooney is President of the United States

DES (I know what the world would be like if Andy Rooney were President).

1: IMG PM (there could be a man with thus and such craggy features sitting at a desk AND Andy Rooney sitting in the Oval Office would look roughly like this)

2: BEL (Presidents gain facility with teleprompters)

Therefore, BEL (If Andy Rooney were President he would gain facility with teleprompters)

IMG PM (that there could be a man with thus and such craggy features glancing at a clear rectangle AND Andy Rooney using a teleprompter would look like this)

3: BEL (Andy Rooney abhors the non-commonsensical)

BEL (Presidents veto bills they abhor)

Therefore (If Andy Rooney were President, he would veto bills in obvious conflict with common sense).

IMG PM (that there could be a red stamp reading ‘VETO’ slamming down upon printed a page AND if I were to see Andy Rooney vetoing a bill it would look roughly like this).

Above we see the generation of beliefs being generated about how the world would be if Andy Rooney were President, with states of sensory imagination serving to represent how the world would look in various relevant situations, were he President. This is what it is, cognitively speaking, to imagine that Andy Rooney is President of the United States. Of course, there is not just one way to imagine that Andy Rooney is President. There are very many ways. The point here is that a single act of imagining that Andy Rooney is President can involve a variety of assertoric states, some of which, when attributed to an imaginer, would involve non-propositional uses of ‘imagine’—e.g., imagining Andy Rooney in the Oval Office; imagining a VETO stamp. These combine with the kinds of beliefs and

inferences I've argued are characteristic of non-imagistic propositional imagination to constitute the mental world of an imager. What at first blush seemed only loosely connected phenomena—sensory imagination and propositional imagination—in many cases have a constitutive relationship: many acts of propositional imagination are partly constructed out of (non-propositional) sensory imagination.

CHAPTER 5:

The Phenomenology of Imaginative Agency

1. *Introduction*

Up to this point, my discussion of imagination—both ‘sensory’ and ‘propositional’—has focused primarily on its representational features. In the case of sensory imagination, I argued that such states have genuine correctness conditions in that they (sometimes incorrectly) represent ways the world looked, or would look. In the last chapter, I argued that propositional imagining consists in generating beliefs in counterfactual or indicative conditionals, based on one’s relevant background knowledge.

I want now to focus on another important aspect of imagination: its relation to the will and to one’s sense of agency. It is a platitude that we willfully control the course of our own imaginings, even if—as we will see—there are times when people imagine things against their wills (and notwithstanding the earlier-discussed cases where people are wrong about what they are imagining). But what is it, phenomenologically speaking, to feel in control of one’s own imaginings? In what does the sense of “freedom” over what we imagine consist? This is a crucial question for understanding the nature of imagination, since the will’s involvement in imagination (as opposed to perception) is, traditionally, one of imagination’s defining features.

So, where earlier chapters challenged the notion that we always know what we are imagining, this chapter looks at how we know *whether* we are imagining. A first thought might be that to feel in control of one’s imaginings is to imagine what one intends to imagine—to have one’s expectations met. But notice that this requires that for each act

of imagining *X* we first generate an intention to imagine *X*. If feeling in control of one's own imaginings requires also feeling in control of one's intentions, a regress threatens. One would need to have an intention to have an intention to imagine *X*, in virtue of which the latter intention seems under one's own control, and so on *ad infinitum*.¹⁴⁰

More importantly, as this chapter will discuss in detail, the empirical literature in psychology reveals a wealth of cases where patients deny that they are the agents of their own imaginings. Is this because they imagine things that they do not intend to imagine? Is it because they are unaware of what they intend? Or is a commonsense appeal to intentions simply inadequate in the face of such data? I argue for the last claim in this chapter, attempting to show that the phenomenological sense we have of being the agents of our thoughts and imaginings is more complex, and of a different nature, than is normally realized.

My method is to look closely at cases (in schizophrenia) where the normal sense of agency attending imagination is disrupted, in hopes of better understanding what the sense of agency consists in normally. I will focus on inner speech as a form of imagination that is abnormal in schizophrenia, moving away from my prior focus on visualization. Nevertheless, imagery in a variety of modalities will be discussed, as will various forms of perceptual phenomenology. I will try to show how the same mechanisms responsible for regulating the sense of agency we have over bodily

¹⁴⁰ And even if feeling in control of one's imaginings does not require feeling in control of one's intentions, one might wonder why a cognitive system would need to generate intentions to imagine thus and such in the first place; why wouldn't whatever resources were needed to generate such an intention be sufficient to generate the imaginative experience itself? Such an intention should not be posited merely to explain how it could be that we feel in control of the act, especially if it is possible to account of the sense of agency in some other way.

movements and visual and auditory perception may also generate the sense we have of being the agents of our own imaginings.

I approach these issues by focusing on a question—called ‘the puzzle of extraneity’—concerning the nature of auditory verbal hallucinations (AVHs) and inserted thoughts in schizophrenia. The puzzle of extraneity asks how it is that one’s own thoughts can seem to be someone else’s. A popular proposal (Campbell, 1999; Gallagher, 2000; Graham and Stephens, 2000; Proust, 2006) has been that the normally fused subjective senses of thought *ownership* and of thought *agency* come apart in schizophrenia. It is the missing ‘sense of agency’ and lingering ‘sense of ownership’ (the latter deriving from the introspective mode of access to the thought) that is usually thought to account for an introspected thought’s being ascribed to another (Campbell, 2002). Yet there is little agreement concerning what the phenomenological ‘sense of thought agency’ amounts to. One goal of this chapter is to show that the phenomenology of thought agency is not what one might pre-reflectively take it to be, and to provide a way of understanding this phenomenology that gives genuine purchase on the puzzle of extraneity.

After rejecting two competing solutions to the puzzle (in section three), I develop a novel solution (sections four and five), drawing on, and amending, the work of Frith (1992, 2005), Campbell (1999), Blakemore (2003, 2005), Jeannerod and Pacherie (2004) and Proust (2006). The central argument of section four is that the popular explanatory paradigm—originating with Frith (1992)—invoking ‘comparator’ mechanisms responsible for calculating the relationship between predicted and actual perceptual (and

cognitive) feedback can be understood in at least two different ways. I discuss these differences, favoring what I call the ‘filter model’ of sensory cancellation. In section five I examine in some detail the potential consequences that deficits in such filtering would have for the phenomenology of thought (and, in particular, of inner speech). Specifically, I try to indicate why such phenomenology would lead to reports both of inserted thoughts and hallucinated ‘voices’. This proposal is supplemented (in section seven) by a discussion of recent research concerning the factors needed to create ‘presence’¹⁴¹ in a virtual environment (VE) (Sanchez-Vives and Slater, 2005). VE research pinpoints some of the cognitive mechanisms underlying the fixity (i.e., the resistance to revision in light of contravening evidence) of delusional belief in a way that supports the present hypothesis, and points the way towards its further testing.

2. The ‘Dual Strands’ of Thought Ownership

Auditory verbal hallucinations and inserted thoughts are two cardinal (so-called ‘first-rank’ (Schneider, 1959)) symptoms of schizophrenia. AVHs are diagnosed when patients report hearing voices in the absence of appropriate stimuli (whether or not the patients believe themselves to be merely hallucinating). Inserted thoughts—typically classified as delusions and not hallucinations—are diagnosed when patients report feeling as though thoughts have been inserted into their minds by a foreign agent. Patients experiencing thought insertion do not take themselves to be hearing the voices of others, but rather feel as though they are subjectively entertaining the thoughts of another person (or even, that someone else is *using* their thoughts to think). Surveying 100

¹⁴¹ ‘Presence’ is a technical term in the VE (virtual environment) literature meant to capture a VE’s ability to instill in the perceiver a sense of actually ‘being there’ in the VE (Sanchez-Vives and Slater, 2005).

schizophrenic patients prone to AVHs, Nayani and David (1996) found that forty-six reported experiencing the distinct phenomenon of thought insertion, where patients' thoughts were reported as 'lacking normal sense of ownership,' seeming 'alien, not their own' (p. 182).

Inserted thoughts, like AVHs, often take the form of verbal comments or commands. Frith (1992, p. 66) quotes one patient's account:

Thoughts are put into my mind like 'Kill God.' It is just like my mind working, but it isn't. They come from this chap, Chris. They are his thoughts.

Another commonly cited example comes from Mellor (1970, p. 17):

I look out the window and I think that the garden looks nice and the grass looks cool, but the thoughts of Eamonn Andrews come into my mind. There are no other thoughts there, only his....He treats my mind like a screen and flashes thoughts onto it like you flash a picture.

Notice that in each case the patient identifies the 'inserted' thoughts as being located in 'my mind'. This leads to a philosophical puzzle: how can someone acknowledge that a thought occurs in her mind while denying that it is her thought? A seemingly inviolable relation is being flouted: isn't any thought in my mind *ipso facto* my own thought? A powerful intuition holds that, when it comes to thoughts, *location determines ownership* (unlike laptops, for instance).

As already remarked, a common way of making sense of such reports is to draw a distinction between the subjective senses of thought ownership and of thought agency, and to suggest that these normally fused aspects of phenomenology come apart in sufferers of inserted thoughts (Campbell, 2002, p. 36). To have a subjective sense of thought ownership without a sense of thought agency is, intuitively, to experience a kind of immediate, (apparently) non-inferential (or 'private') *access* to the content of a

thought, while at the same time lacking any phenomenological sense of being causally responsible for the generation of that thought (Campbell, 1999).

Once we unpack the idea in this way, however, it becomes evident that normal subjects also experience this ‘split’ from time to time, such as when one finds oneself with a song fragment repeating in one’s mind, or simply engages mechanistically in a routine activity (an activity which nevertheless involves conscious thought). These quotidian experiences do not lead us to ascribe the agency or ownership of such thoughts to an external agent. One might also question whether most thought really is accompanied by a sense of causal agency or ‘effort’ at all (Strawson, 2003). Perhaps when solving complex problems or negotiating new situations, we may have a robust sense of being the causal agents of our thoughts and decisions—of controlling and shaping the course of our thinking—however it is far from clear that at other times there is anything properly called a ‘phenomenology of agency’ attending thought (at least, not if we think of this phenomenology as an effortful sense one has of *trying* to think in certain ways). The unsupported assumption that there is a clearly understood phenomenology or ‘emotion’ of thought agency—one that can explain reports of thought insertion by its absence—is a pervasive weakness in the literature on this subject.¹⁴²

Thus, simply appealing to a dissociation between ‘two strands’ in the ordinary conception of thought does not take us very far in answering the puzzle of extraneity. Two challenges remain: first, to give a clear account of the nature and causal origin of the agency-phenomenology that becomes abnormal in thought insertion. This account should be consistent with the fact that we often think without experiencing any sense of

¹⁴² Cf. Frith (1992, p. 81), and, more recently, Jones and Fernyhough (2006) and Proust (2006).

effort or of an exertion of cognitive force. Second, to explain the phenomenology of thought insertion in a way that reveals why schizophrenic patients not only fail to feel causally responsible for their thoughts, but also feel that their thoughts *belong to someone else* (thereby explaining the *extraneity* attending their phenomenology). To meet these challenges is to solve the puzzle of extraneity.¹⁴³

3. Alternative proposals

Before going on to develop a positive solution to the puzzle of extraneity, it will help to consider some rival accounts—both in order to appreciate the depth of the puzzle, and to reveal a connection between auditory hallucinations and inserted thoughts that will be important in later sections.

3.1 The Sensory Resemblance Thesis

Perhaps the first idea that comes to mind in trying to answer the puzzle of extraneity is that patients may ascribe the agency or ownership of both auditory verbal hallucinations and inserted thoughts to others simply because what it is like to hear voices and have inserted thoughts is similar to what it is like to actually hear others speak—in essence, like causes are inferred from like effects (Slade and Bentall, 1988; Kinsbourne, 1990).

Call this the Sensory Resemblance Thesis (SRT). SRT may indeed solve the puzzle of

¹⁴³ While the extreme oddity of schizophrenic symptoms may tempt one in the direction of skepticism concerning psychological, phenomenological, or cognitive explanations (it may seem that there is no longer a stable enough belief-desire economy in schizophrenia to ground a cognitive theory), this outlook is unsupported by the empirical literature. Maher (1974, p. 99) cites a wide range of studies indicating that ‘when adequate controls for variables such as education, verbal knowledge, etc. are introduced, there is no basis in evidence for the belief that schizophrenic patients differ from normals in their vulnerability to any type of [logical] error.’ And, indeed, many of those afflicted with schizophrenia carry on successfully in day-to-day reasoning tasks, the peculiarities of their experience revealed mainly in the reports of their hallucinated voices or inserted thoughts, the content of which is typically ‘fairly narrow’ (Frith, 1992, p. 79).

extraneity for some auditory hallucinations. However, closer attention to the reported phenomenology of auditory hallucinations and inserted thoughts shows SRT to be an insufficient solution to the puzzle of extraneity for most such episodes. To see why, we have to look more closely at the relationship between AVHs and inserted thoughts.

Given that inserted thoughts can take the form of verbal commands and commentary, and are experienced as unwilled by the subject, how are they different, phenomenologically, from auditory hallucinations? One natural suggestion is that auditory hallucinations are simply more vivid, or possess richer sensory character, making them more likely to be mistaken for cases of actually hearing someone speak, resembling less the mere verbal-imagery-involving thought (or ‘inner speech’) that most people engage in from time to time (as when one debates an imagined opponent in an inner dialog). However, some patients report experiencing *voices* (as opposed to inserted thoughts) that lack any auditory or sensory component (Graham and Stephens, 2000, p. 99-100). How should *these* experiences be distinguished from thought insertion?

The correct answer may be that they should *not* be so distinguished. Graham and Stephens take this approach, giving reasons for thinking that auditory hallucinations and inserted thoughts should be treated as kindred phenomena, the distinction arising partly out of patients’ difficulties in reporting the nature of their experiences, and partly out of differences in the severity of a single disorder. In support of this idea, they discuss the work of Junginger and Frame (1985), who asked 26 hallucinating patients to rate their voices on a 1-10 scale, where 1 indicated that the voice seemed to exist entirely ‘inside the head’ and 10 indicated the voice seemed clearly to come from ‘outside the head’ (i.e., was like hearing someone else speak). Their results were mixed: 11 subjects (41%)

rated their voices at 5 or below, while 15 (59%) rated them at 6 to 10. A minority of patients reported that hearing voices was more like ‘hearing yourself think’ than like hearing somebody talk. Nayani and David’s (1996) survey supports these findings: 49% of patients reported hearing voices ‘through their ears as external stimuli’, 38% reported hearing them ‘in internal space,’ and 12% heard them in both loci variably (p. 180).

Such reports weigh in favor of two conclusions: first, many episodes classified as auditory hallucinations (and all inserted thoughts) are not, for their subjects, similar to experiences of actually hearing someone speak (normally perceived voices do not seem to come from inside the head, or to lack sensory character). Therefore, the agency responsible for such episodes must be outwardly ascribed for some reason other than what the SRT posits.

Second, we have reason to think that some reports of voices (those towards the lower end of Junginger and Frame’s scale) may be reports of the very same phenomenon that others report as inserted thoughts. Just as one might, in forming a visual image of one’s childhood home, say ‘I can see it now,’ (even if having visual imagery is only a little like seeing), so might one who experiences unusual inner speech say that he ‘hears’ various comments and commands. In short, patients may be using normal modes of expression in a rough-and-ready way, in order to communicate the nature of experiences that do not fall neatly into any preexisting category.

These reflections are supported by the pioneering work of Eugen Bleuler (1911/1950, p.110), who found that his schizophrenic patients sometimes described their voices as ‘soundless’ and as like ‘vivid thoughts.’ Bleuler notes that ‘even intelligent patients are not always sure that they are actually hearing the voices or whether they are

only compelled to think them' (*ibid.*). Though Bleuler's work predates the auditory hallucination/thought insertion distinction, something very much like it is present in his distinction between 'hallucinations of perceptions' and 'hallucinations of conceptions.' These are 'differentiated by patients' accordingly as the voices appear to 'come from the outside' (hallucinations of perceptions) or, alternatively, seem 'projected into their own bodies' having 'hardly any sensory components' (hallucinations of conceptions) (1911/1950, p. 111). Here, too, the key role played by the patient in 'differentiating' one phenomenon from the other should encourage us to leave open the possibility that the very distinction arises out of alternate ways of describing a single unusual phenomenon. Also, given that the mechanisms responsible for more 'vivid' auditory hallucinations likely overlap with those responsible for inserted thoughts and 'soundless' voices, we should also leave open the possibility that whatever solves the puzzle of extraneity for the latter may also solve it for the former.

3.2 *Graham and Stephens' 'Intentionally Inexplicable' Thoughts*

Graham and Stephens take a different approach to the puzzle of extraneity, arguing that it is the 'apparent intelligence' of inserted thoughts and hallucinated voices that provides the 'experiential or epistemic basis' for the attribution of their agency to another (2000, p. 174). Of course, for the apparent intelligence of a thought to lead one to other-ascribe its agency, one must *first* not take oneself to be the agent of that thought (unless one is utterly convinced of one's own unintelligence). According to Graham and Stephens, a person's initial *lack* of a sense of agency results from a given thought or mental episode not being 'explicable in terms of his conception of what he believes and desires' (2000, p.

165). On their view, people continually monitor the content of their own thoughts and, when thoughts are detected which do not ‘fit’ their other beliefs and desires (being, in Graham and Stephens’ term, ‘intentionally inexplicable’), these thoughts acquire an alien phenomenology. Thus, for Graham and Stephens, there is nothing *phenomenologically* foreign about inserted thoughts or hallucinated voices *prior to* their detection as being out of keeping with one’s beliefs and desires, allowing them to *explain* the phenomenology of non-self-agency in non-phenomenological terms.

If these ill-fitting thoughts or voices then *also* seem intelligently structured, an external ascription of the agency responsible for them will occur as a sort of inference to best explanation. In essence a kind of ‘argument from design’ is applied, roughly as follows:

P1) These thoughts I am entertaining require an active intelligence for their creation (they are not mere ‘parrotings’ or repetitions of earlier perceived events).

P2) I am not the active intelligence responsible for them (due to their lack of ‘fit’ with the self-narrative).

C) Someone else is creating/thinking these thoughts.

Often, the voices and inserted thoughts experienced by patients are accusatory and belligerent. Such voices and thoughts—apparently expressive of an antagonistic ‘intelligence’ of some kind—clearly oppose one’s larger goals and self-conception, and provide the key examples for Graham and Stephens’ account of the sense of thought-agency. However, as Graham and Stephens themselves note, many voices and inserted thoughts are generally positive in their content, their messages consistent with the larger plans and desires of their subjects. Bleuler notes that ‘besides their persecutors, the

patients often hear the voice of some protector' (1911/1950, p. 98). He also reports that, occasionally, hallucinatory voices 'represent sound criticism of [the patient's] delusional thoughts and pathological drives' and therefore should be generally consistent with whatever non-delusional thoughts he may still have (*ibid.*). Graham and Stevens also discuss Chapman and Chapman's (1998) report of a young woman who persistently heard the voices of her deceased father and grandmother. She regarded these voices as benevolent and generally beneficial to the achievement of her practical goals (such as when they advised her on the purchase of a new car) (2000, p. 169-170).¹⁴⁴ Such cases show that one's inability to intentionally explain a thought is not necessary in order for that thought to have an element of extraneity.

Nor is one's inability to intentionally explain a series of intelligently structured thoughts sufficient to make them seem to be owned or produced by another agent. Few people, before the onset of mental illness, believe in the existence of (actually) inserted thoughts and (actually) perceived voices of invisible agents, so any intelligently structured belief in such things will itself (at least initially) be radically out of keeping with one's other beliefs and desires. Graham and Stephens' theory should predict that the very thought that one has had thoughts inserted into one's mind (or that one is hearing the voice of an invisible agent) will itself seem not to be one's own thought. And yet these (delusional) thoughts are precisely the ones that *are* self-ascribed by schizophrenic patients, and are the ones theorists must explain. It is plainly self-defeating for a subject to other-ascribe the agency of her own thoughts for the sole purpose of preserving a

¹⁴⁴ In light of such cases, Graham and Stephens suggest 'it may be necessary to tell different stories about why the subject finds her thoughts intentionally inexplicable in various cases' (2000, p. 171). It is not clear that this explanatory blank check can be cashed. If showing a thought to be intentionally inexplicable requires considerable explanation from case to case, the overall form of explanation will inevitably seem *ad hoc*.

consistent self-narrative, for such a claim immediately turns one into the protagonist of a very different sort of story.

It is worth noting that Graham and Stephens' theory may still explain the lack of a sense of thought-agency that attends certain *non-pathological* thought episodes. Take, for instance, the novelist who feels she must write in order to discover what her characters will do, or the songwriter who feels he is merely a medium through which some other agency is expressed. Unable to gain conscious access to the aspects of their psychologies responsible for their artistic creations (i.e., unable to intentionally explain their acts), artists—and highly-trained individuals in many other fields—occasionally feel as though they are not the agents of their thoughts and actions. However, this mere lack of an intellectual understanding of how one accomplishes one's acts is a far cry from the pathological phenomenology of other-agency that characterizes schizophrenic experience. The obvious difference remains that this 'artist's extraneity' rarely reaches the point where a delusional belief is formed that *someone else* actually is the agent of one's thoughts. This is a difference that calls for a different explanation in terms of the cognitive mechanisms responsible for the respective kinds of cases. Nevertheless, it is helpful to have Graham and Stephens' account at hand to explain cases where reports superficially similar to those of schizophrenic patients are made by normal subjects, if only to more clearly distinguish them from the cases of thought insertion to be discussed.

To briefly consider a further option, one can simply posit a certain phenomenological 'my-ness' (perhaps taken for granted in normal thinking) which is lacking in the case of inserted thoughts and 'soundless' auditory hallucinations, a palpable 'otherness' existing in its stead. Yet, this sort of phenomenological 'answer' to

the puzzle of extraneity does little more than restate the problem and assure us that the patients are not lying—it tells us that *it really does seem to them* that their thoughts are controlled (or owned) by another. Though doubtless true, this does nothing to explain *what it is* for their thoughts to lack the normal phenomenological character of ‘my-ness’. Just what is this ‘my-ness’, other than a label for the problem? And why think that we can easily form a clear notion of what it would be like for it to go missing? Those who have never had such experiences may be no better suited to imagining them than the blind are to imagining sight. So, the question of the *character* of the delusional experience remains open.

Nor do phenomenological answers explain *how* such phenomenology could come to exist in an otherwise rational being. We still need to identify the mechanisms in normal humans whose malfunction can lead to these specific symptoms. To avoid being *ad hoc*, or a mere restatement of the problem, the mechanisms cited in the explanation should also help explain phenomena outside of the deficits they are here called on to explain. A last criticism of purely phenomenological approaches concerns their appeal to normally unnoticed or ‘overlooked’ elements of phenomenology (the ‘sense of agency’ or ‘my-ness’ of thoughts). Given that a creature’s phenomenology is just a matter of how things seem to that creature, the idea of a ‘taken for granted’ or unnoticed aspect of phenomenology is obviously problematic.

4. Sensory-Feedback Approaches

While phenomenological approaches face problems, purely neurological approaches have shortcomings as well. It has been known since the late 1950s that schizophrenic

symptoms are correlated with increased levels of dopamine absorption in the brain (indeed, this is one of the main reasons researchers are confident the disease has a unitary biological basis)—hence the common prescription of dopamine-blockers as antipsychotic drug treatments. However, the vocabulary of ‘dopamine’ and ‘neurotransmitters’, etc., is too distant in terms of causal/theoretical links from that of psychology to provide the kind of explanatory purchase on schizophrenia that would allow for a more nuanced (and perhaps even curative) treatment. For all current researchers know, increased dopamine absorption may be an effect of the underlying disorder, and not its cause. Hence the move theoretically ‘upward’ to the cognitive-modular level of explanation.¹⁴⁵

At this level of explanation, the oddity of schizophrenic reports is attributed to correspondingly unusual (and highly variable) experiences. These may be experiences normal perceivers do not know what it is like to have, yet which nevertheless arise from commonly shared cognitive mechanisms and processes. Importantly, the mechanisms and processes appealed to in such explanations are ones that have a theoretical life outside of the peculiar puzzle they are here called on to explain.

4.1 Frith, Efference Copies, and the Attenuation of Sensory Input

Frith (1992) was one of the first to develop a detailed cognitive analysis of this kind—a ‘box and arrow’ analysis pitched at the level of functionally specialized cognitive mechanisms, which map (very) roughly onto various regions of the brain. The operation of such mechanisms often involves processes and informational (or representational) states that are themselves not available to consciousness; their existence is typically

¹⁴⁵ For an extended discussion of the issue of levels of explanation in schizophrenia by a cognitive scientist, see Frith (1992, p. 25-29).

posited on the basis of comparative studies in psychology and neurology. In essence, a level of explanation is sought which may provide more explanatory resources than those available at the commonsense phenomenological level, but which is also functionally *near enough* the level of psychological explanation for its (mal)functioning to have clear implications for the specific kinds of abnormalities reported by schizophrenic patients.

A guiding insight of Frith's (1992, 2005) account of the cognitive basis of schizophrenic delusions and hallucinations is that the prediction (and subsequent monitoring) of the sensory consequences of an action plays a key role in allowing the subject to distinguish between self and other-generated changes in sensory input. This link between perception and the phenomenology of agency—accepted also by Blakemore (2005) and Jeannerod and Pacherie (2004), among others—stems from the recognition of a problem faced by the early perceptual system. Moving one's eyes causes changes in the signal detected by the retina, but there is nothing in this changing signal *per se* that allows the perceptual system to determine whether the change is caused by one's own willful eye and head movements, or by an independent event perceived in the environment. Sperry's (1950) (and, concurrently, von Holst and Mittelstaedt's (1950/1973)) well-known solution to this problem (drawing on earlier work by Helmholtz (1866)) was to propose that motor commands resulting in eye movements release an 'excitation pattern' that includes a signal (often called 'efference copy') which tells the perceptual system to anticipate certain kinds of changes in incoming retinal information (the incoming sensory information is called 'reafference'). Sperry calls this 'adjustment factor' allowing for the anticipation of certain kinds of input a 'corollary discharge', and suggests it could 'aid in maintaining stability of the visual field under

normal conditions during the onset of sudden eye, head, and body movement' (Sperry, 1950, p.488).

A simple experiment reveals the power of this proposal to shed light on the phenomenology of agency. With one eye closed (or a hand over one eye), push lightly several times on side of the eyelid of your open eye. Your visual image should appear to 'move' or 'wobble'. However, if one creates a similar retinal change by simply moving one's head in the proper direction, no such 'wobbling' of the visual image occurs.¹⁴⁶

Why?

Sperry's explanation is that applying pressure to the eyeball with one's finger results in changes of the retinal image in the absence of the normally triggered efference copy signal (and resultant 'corollary discharge').¹⁴⁷ The efference copy signal is not triggered in this case because the change in one's retinal display results, atypically, from finger pressure, not the eye, head, and body movements that typically result in retinal changes and accompanying efference copy signals. The corollary discharge triggered by these signals filters out potentially distracting or irrelevant elements of the incoming information so as to result in a stabilized visual image. Without the stabilizing effect of the corollary discharge, the visual phenomenology is as of the room itself moving—and, indeed, if it were not for contrasting information simultaneously registered by other

¹⁴⁶ A similar phenomenon is present in the opposite direction as well: when eye muscles are partially paralyzed with curare (Brindley and Merton, 1960) and the subject told to move his eyes, the subject reports the visual appearance as of the room moving in the direction in which the eye movement would have occurred; though the retinal display has remained *exactly* the same, the entire room seems visually to shift in the direction of the merely intended movement, since, under normal conditions, that sort of retinal display, when filtered in accordance with an efference copy signal, would correlate with the room's actually moving in the direction of the shifted gaze.

¹⁴⁷ Some theorists (Gallagher, 2000; Campbell, 1999) identify corollary discharge with efference copy, while others (Campbell and Pettigrew, 2004) insist that the two are distinct phenomena—see fn. 8 for elaboration on this point.

sensory and proprioceptive systems (and the knowledge that one has simply pushed one's eye), it would indeed seem as though the environment itself were moving (as in an earthquake), and not simply one's visual field. Thus, the same structure of changes on the retina can result in very different phenomenologies, depending on whether the efference copy/corollary discharge mechanisms are active. One phenomenology (the stable visual field) will naturally be associated with one's own willful actions, while the other will be associated with perceptual changes caused by events outside of one's agency.

This 'filtering' of retinal information through the efference copy/corollary discharge process is just one instance of a general phenomenon, whereby the sensory consequences of self-initiated actions are 'automatically' (i.e. pre-consciously) attenuated (Blakemore, 2005). Studies by Blakemore *et al.* (1998, 2003), Ford and Mathalon (2005), Shergill *et al.* (2005), and Lindner *et al.* (2005) indicate that schizophrenic patients are often unable to achieve this attenuation, across multiple sense modalities (including proprioception). For instance, normal subjects are generally unable to tickle themselves, their sensitivity to their own touch being greatly attenuated (Weiskrantz, Elliot, and Darlington, 1971). Frith (2005) notes that this attenuation of self-generated sensation can also be observed at the physiological level, as activity in the somatosensory cortex is much reduced when tactile stimulation is self-applied. Schizophrenic patients do not show this attenuation, rating their sensitivity to self-touch as high as the same touch applied by someone else (Blakemore *et al.*, 1998). Also, brain imaging reveals that responses to sound in auditory cortex are normally attenuated when a person is speaking. This attenuation is not found in many schizophrenic patients (especially those who report

auditory hallucinations) (Ford and Mathalon, 2005). Ford and Mathalon (2004) also found that, while inner speech reduces responsiveness of auditory cortex in normal subjects, schizophrenic patients lacked this attenuation during inner speech. In the case of vision, Lindner *et al.* (2005) found that schizophrenic patients with delusions of influence were more likely to attribute the sensory consequences of their own eye movements to the environment rather than to themselves.

4.2. Two Ways of Understanding Corollary Discharge

A potentially powerful explanatory paradigm invoking efference copies and reafferences—sometimes called the ‘central monitoring’ or ‘comparator’ theory—has been developed and adopted in one form or another by Frith (1992, 2005), Blakemore *et al.* (2002, 2003), Shergill *et al.* (2005), Linder *et al.* (2005), Jeannerod and Pacherie (2004)¹⁴⁸, Campbell (1999), and others. These theorists all make explicit appeal to a ‘comparison’ that is made between an efference-copy-elicited ‘prediction’ (or ‘corollary discharge’) and reafferent (sensory or cognitive) input. Insofar as the two signals ‘match’, the incoming signal is said to be ‘cancelled out’ (resulting in its overall attenuation) (Blakemore *et al.*, 2002, p. 238).¹⁴⁹ However, this notion of a modality-

¹⁴⁸ Jeannerod and Pacherie develop a simulation-based theory importantly different from Frith’s theory in how it treats ‘covert’ actions, but nevertheless claim their theory to be ‘largely complementary’ to Frith’s paradigm (2004, p. 125).

¹⁴⁹ There is no common understanding in the literature of what ‘efference copy’ and ‘corollary discharge’ refer to. Campbell (1999) and Grush (2004) identify efference copy with corollary discharge, while Blakemore *et al.* (2003, see esp. Fig. 1, p. 1059)—drawing on Miall *et al.* (1993)—use ‘efference copy’ to refer to the copy of the motor command that is sent to a later system that in turn generates a prediction of the incoming state, this prediction being called a ‘corollary discharge’. Thomas Campbell and John Pettigrew (2004) agree that there is a distinction between efference copy and corollary discharge, but use the terms in precisely the opposite manner, suggesting that efference copy is the sensory prediction compared with input, and corollary discharge the copy of the motor command that results in the generation of the prediction.

specific *comparison* is not explicitly present in Sperry or von Holst's classical appeal to efference-copy-driven attenuations and cancellations of sensory feedback; rather, it grows mainly out of the work of Miall *et al.* (1993) and Wolpert *et al.* (1998) on 'forward models', hypothetically used by the motor system to adjust and control limb movements.

In brief, a 'forward model' of the kind posited by Miall *et al.* and Wolpert *et al.*—and expanded on in important ways by Grush (2004)—generates predictions of sensory feedback, based on efference copies of commands issued by the motor system. Some such predictive mechanisms seem necessary to explain corrective arm adjustments made in grasping tasks that occur too quickly (200-300ms) to result from the visual or proprioceptive monitoring of sensory feedback (Miall *et al.*, 1993, p. 205; 1998, p. 343; Wolpert and Ghahramani, 2000). If motor commands to engage in arm movement involve the triggering of an efference copy of the command, which itself generates a 'prediction' of the final state (perhaps by enacting an 'offline' emulation of the action (Grush, 2004)), this prediction can instantaneously be 'compared' with the resultant state to determine whether the desired movement was accomplished. Any discrepancies between the forward output model's prediction and reafferent feedback detected by the comparator mechanism will cue the organism to adjust its behavior. If, however, the prediction and feedback 'match', the reafferent signal is cancelled out, making us 'largely unaware of the sensory feedback about the actual state of our motor system as long as our intentions have been achieved' (Blakemore *et al.*, 2002, p. 238). Notably, Miall *et al.* (1993), Wolpert *et al.* (1998), and Wolpert and Ghahramani (2000) only explicitly

I adopt Blakemore's *et al.* (2003) use: efference copy is a parallel 'copy' of a motor command, while corollary discharge is a prediction (or, in my case, a filter command) of the sensory consequences of the action, made based on the (efference) copy of the motor command.

discuss and defend the operation of such models with regard to *limb movements*, and do not link their work to the classical models of Sperry and von Holst (Wolpert and Ghahramani (2000, p. 1217) do *anticipate*, however, that forward models ‘will be found to underlie the control of motor systems as diverse as the eye, arm, speech, posture, balance and locomotion.’).

Appropriately, Blakemore *et al.* (2003) draw on Miall *et al.* (1993) and Wolpert *et al.* (1998) to explain cases where schizophrenic patients do not feel in control of their own bodily movements. Yet, in attempting to extend such findings to explain auditory hallucinations and inserted thoughts, Frith (1992, 2005), Blakemore *et al.* (2003), Jones and Fernyhough (2006), and Blakemore (2005) run together discussions of Miall/Wolpert-type forward output models with Sperry and von Holst’s more general account of corollary discharge, giving the impression that forward output models are the only (or at least best) way of implementing Sperry and von Holst’s theories for the auditory and visual systems.

But there is another way of understanding how corollary discharge and sensory cancellation may take place in the visual and auditory systems—one which has some advantages when it comes to explaining inserted thoughts and AVHs. Before offering that account, I will describe in some more detail how the sort of forward output model typically discussed in the literature would apply to the visual system, to clarify the contrast between it and the theory I will offer.¹⁵⁰

¹⁵⁰ As I note below, some (but not all) accounts of forward output models (e.g., Wolpert and Ghahramani (2002)) are sufficiently abstract as to be neutral between the two accounts of sensory cancellation I here describe.

Suppose, for simplicity, that the retina is a two dimensional square grid containing nine cells, each of which can take the values ON and OFF. We can refer to the cells by their coordinates—the lowest left cell being (1,1), the upper right cell being (3,3), and so on. The present state of the retina can thus be given by a list indicating whether each coordinate pair is ON or OFF. According to most versions of the forward output model, the motor and/or visual system uses an efference copy of a motor command (e.g., look to the right) to calculate a prediction of the sensory feedback that will be received. This calculation requires as input both the command that is sent and an initial state of the retina (in the case of vision). Suppose that the values of the nine-cell “retina” in its initial state (S1) are ON for cells (2,1), (2,2) and (2,3), and of OFF for the rest. An efference copy of the command “look to the right” generates a prediction state (P1), itself a nine-celled grid with (let us suppose) the values ON for (1,1), (1,2) and (1,3) and OFF for the rest.¹⁵¹ P1 is then ‘compared’ with actual sensory input (S2). If S2 is in fact ON (1,1), (1,2) and (1,3) and OFF for the rest, then the prediction and subsequent input match, and the incoming information is cancelled (i.e. prevented from carrying forward to further processing). Note that, for this cancellation to occur, a separate ‘comparator’ mechanism (or algorithm) is needed that, in effect, makes the judgment that P1 does (or does not) match S2.

This is, I take it, how a forward output model involving predictions and comparisons could function in the visual system (it is, in essence, what Grush (2004, p. 189) proposes). Yet it seems that the same sensory cancellation could occur in a simpler

¹⁵¹ Note that prediction state (P1) could potentially play its role while being a coarser grained representation than the state to which it is compared—i.e., while having a determinate value for only *some* coordinate pairs. Nothing in my discussion will hinge on this point.

way. We know that information on the retina changes in law-like ways each time one looks to the right. This means that, given an initial state of the retina, and a copy of a motor command to look to the right, the visual system could simply predict the needed cancellation without ever generating a simulation of the input itself (*ala* P1 above) to be compared with sensory input. That is, given S1 (ON for (2,1), (2,2) and (2,3) and off for the rest), together with a command to look to the right, a well-trained (or adapted) visual system could simply issue another command to turn OFF cells (1,1), (1,2) and (1,3) in S2. In effect, for *every* case of looking to the right, the system could apply to each subsequent state the rule ‘turn OFF each cell to the left of any cell that is ON in present state,’ using each present state to determine which specific cells those should be from case to case.

Mathematically, this model accomplishes the same cancellation as the forward output model described above. It also can be said to feature a kind of prediction—the negative, input-cancelling state that is issued can be seen as an anticipation that a certain kind of eye movement will occur (and, therefore, that a specific kind of attenuation is needed). Yet there are important differences in how predictions and cancellations are accomplished by the two models. On the model I am suggesting (the ‘filter model’), there is no need for a mechanism that *compares* the predictive signal to sensory input. The filter signal may simply be *applied* to S2 without need for a judgment of its relation to S2, just as a sequence of delete commands could remove every third frame of a film without requiring any comparison between the deletion signals themselves and the film.

To briefly expand on this last metaphor, suppose that in the editing of a film a particular temporal sequence of delete commands (D1) is needed following any frame containing only red (because certain patterns of unnecessary frames invariably follow

completely red frames). The triggering of D1 could be thought of as a prediction or anticipation that these unnecessary frames will be occurring. But there are not, literally, any states in the editing system or film that are *compared* with the deleted frames during this ‘attenuation’.

Removing the need for a comparison allows the system to apply one less ‘rule’ during perception. Recall that, on the comparator or forward output model described above, the cognitive system takes S1 and a motor command as input, and applies a rule (an algorithm) to generate P1 (the prediction). Then an additional rule must be applied that governs how to change or attenuate S2 given P1 (this is the rule or algorithm that, in effect, ‘compares’ P1 to S2). On the filter model, the cognitive system takes S1 and a motor command as input and simply applies an appropriate negative algorithm to S2.¹⁵²

Note also that the filter model does not require that the ‘predictive’ signal itself be a quasi-visual state similar in kind to S1 and S2. The ‘prediction’ could simply be a negative signal that, like the delete command of the film editing system, has the effect of turning off certain targeted cells of S2, without itself representing the presence or absence of an object at some point in space (this is because it acts only to attenuate or delete certain parts of S2, while ‘saying’ nothing one way or another about the rest of the visual grid). The point, in more intuitive terms, is that we need not form an image in order to attenuate or ‘filter’ one.

Defenders of comparator or forward output models may reply that their models require nothing more than the kind of filtering I have suggested, and that the cancellation

¹⁵² It is assumed on both models that, given S1 and a copy of the motor command, the cognitive system has the resources to generate an appropriate ‘predictive’ signal—be it a negative ‘filtering’ signal or a positive one sent to a comparator. How the system initially learns to generate this appropriate signal is an interesting further question—see Wolpert and Gharamani (2000, p. 1216-1217) for discussion.

I have described can still be thought of as a sort of comparison. I find it odd to think of the described filtering as involving comparison, but the commonsense terminology used to describe such processing is ultimately unimportant. What I hope to have made clear is that there are (at minimum) two importantly different ways that sensory cancellation may occur. If both are ultimately consistent with a suitably abstract understanding of the forward output model paradigm, then at least some clarity concerning possibilities for its implementation has been achieved.

Many discussions of corollary discharge are simply too abstract to make it obvious which of the two described pictures is favored.¹⁵³ Some, however, clearly favor the initial picture, involving a comparison between a P1 state in expressed visual coordinates and S2. Grush is explicit concerning the quasi-perceptual nature of the predictive states triggered by efference copies, arguing that ‘emulators’ (his term for ‘forward models’) take motor commands as input and produce ‘an output signal...identical or similar to the feedback signal produced by the plant’¹⁵⁴ (2004, p. 379). On his theory, visual imagery itself is “‘mock” input generated from the operation of an internal emulator’ offline (2004, p. 390). A commitment to predictive and input states being of the same essential kind is also explicit in Miall *et al.* (1993, p. 210): ‘Because the internal models must predict the expected outcome of movement in a form suitable to compare with the actual outcome, these internal models would operate in visual or egocentric coordinates.’

¹⁵³ For instance, Ford and Mathalon (2004, p. 37) note that ‘efference copy works to suppress perception when it results from self-generated action,’ allowing ‘an automatic distinction between internally-generated and externally-generated percepts.’ They go on to describe a range of experiments implicating corollary discharge dysfunction in schizophrenic patients, while saying nothing that would favor one of the models I have discussed over the other.

¹⁵⁴ ‘Plant’ is a control-theory term for the system being controlled, such as the body or the eye.

Recognizing that sensory cancellation via prediction does not *require* a comparison between similar kinds of states has several important consequences. First, as already noted, it means that we need not posit a separate comparator mechanism that calculates the difference between two similar states. Second, because it does not require that a quasi-perceptual (for vision, an ‘imagistic’) state be generated in order to accomplish sensory attenuation, it suggests a difference in where, neurologically, researchers should look when trying to locate the source of sensory attenuation disorders. In the case of vision, one would not necessarily investigate areas responsible for generating visual imagery, since, *pace* Grush, corollary discharge may not draw on imagery at all.

Of course, it remains an open empirical question whether the filtering hypothesis or the comparator hypothesis as I have described it is in fact correct. Both proposals are compatible with recent work that has explored efference-copy driven sensory cancellation in the visual system and its deficits in schizophrenic patients (Lindner *et al.*, 2005). They are also both compatible with Sperry and von Holst’s classical work on the visual system.¹⁵⁵

However, accepting the filtering picture in lieu of the comparator model has the important advantage of being extendable to an account of inserted thoughts. According to the comparator model, where one’s sense of agency is grounded in comparisons between similar kinds of states, for a *thought* to lack a sense of agency, there would need

¹⁵⁵ Sperry (1950, p. 488) posits ‘anticipatory adjustments’, made only with regard to the direction and speed of each eye movement. von Holst (1950, p. 150) suggests thinking of efference copies as ‘positive’ anticipatory signals that, when meeting with an appropriate ‘negative’ incoming sensory signal at some early level of processing, cancel out the positive signal—precisely as 4 cancels out -4 when the two are added—preventing the input from continuing into higher processing areas. Neither makes an explicit appeal to ‘comparisons’ taking place.

to be a prediction of that thought (i.e., an intention to think that *X*) that is compared with the actual thought (e.g., the thought that *X*) (Campbell (1999) develops a theory along these lines). A number of theorists have questioned whether we can plausibly suppose that there really are intentions to think *that X* corresponding to each actual thought *that X* (Gallagher, 2004). What would be the purpose of such a mechanism, given that one's cognitive system is not normally faced with the problem of distinguishing between other and self-generated thoughts?

Campbell hypothesizes that such mechanisms might allow one to continually monitor the semantic appropriateness of one's thoughts, allowing one to maintain a coherent train of thought. But, as Gallagher (2004, p. 12-13) argues, if the thoughts in question are already conscious, why isn't this enough to ensure that they are appropriately monitored? Indeed, why think that a sub-conscious 'thought-checking' mechanism would be in a better position to judge the semantic appropriateness of a thought than the conscious subject himself? Also, the envisioned role of these thought-predictions—keeping one's thought 'on track'—is quite different than verifying that the thoughts are self-generated, so it remains unclear how defects in this mechanism could lead to one's thoughts seeming to be generated by someone else. The filter model sidesteps these worries by removing the need to suppose we continually (albeit sub-consciously) predict what we are about to think. This is because the state accomplishing cancellation is not of the same basic kind as the one being cancelled. In the next section I discuss how filtering, applied to the internal reafferences generated during inner speech, may imbue inner speech with a phenomenology of agency.

5. The Filtering Hypothesis and Inner Speech

The potential implications of the filter model for experiences of auditory hallucinations and inserted thoughts in schizophrenia can now be sketched: if inner speech is a kind of mental action involving internally generated ‘reafferences’ of its own—reafferences which may not be properly filtered—one might anticipate phenomenologies of the kind reported as inserted thoughts and ‘soundless’ voices. Just as normal perceivers can ‘nudge’ their eyeballs to achieve a certain ‘fractured’ phenomenology (forcing one to cite ‘movement’ in one’s visual field, as opposed to the external world), so too may deficits in the efference copy mechanisms of schizophrenic patients lead to a fracturing of the phenomenology attending inner speech.

By a ‘fracturing’ of phenomenology, I mean the presence of a phenomenology normally associated with external changes in the environment that comes to seem as though it is only registering internal changes, due in part to its lack of coherence with other sensory and proprioceptive information. In the visual case, the wobbling visual field is a visual phenomenology normally associated with the entire environment moving (as in an earthquake) that comes to seem as though it is only registering internal changes, due its lack of coherence with the information available to other sense modalities (note: the relevant fracturing is in one’s *overall* phenomenology, as it results from conflicting messages from different sense modalities).

Revealing theoretical and empirical evidence for how unfiltered inner speech could result in a similarly fractured phenomenology—and linking this fractured phenomenology to the phenomenology of thought agency—is the central concern of what follows.

5.1 Imagery, Covert Actions, and Internal Reafferences

The critical task is now to show how the abnormal inner speech of schizophrenic patients can be explained by analogy to the filtering model postulated for the visual system.

There is a *prima facie* difficulty with the analogy: inner speech does not involve the input and processing of new information from one's environment, and thus would appear to involve no reafferent inputs of the kind that must be filtered or attenuated. This problem has led some (Jeannerod and Pacherie, 2004; Proust, 2006; Grush, 2004) to search for ways to assimilate the mechanisms at play in perceptual and motor activity to those at work in 'covert' actions (such as action planning and imagination) as well.

While the notion of a 'reafference' is traditionally only applied to perceptual experience, Proust suggests that, from the standpoint of simulation theory, we have reason to posit something equivalent in the case of merely simulated or imagined actions. For Proust,

There is no fundamental difference whether control applies to external or to internal actions. In both cases, the brain uses its own internal states and *stored reafferences* to simulate and regulate its own processes (2006, 4.1, emphasis added).

Jeannerod and Pacherie (2004, p. 125) work from a similar hypothesis, suggesting that 'actions, whether or not they come to execution, are centrally simulated by the neural network, and...this simulation is the basis for action recognition and attribution.' A key part of their view, as with Proust's, is that actions are simulated 'whether or not they come to execution'—thus, they are simulated in the absence of input from the environment.

According to these theorists (and Grush (2004)), an internally generated prediction state, expressed in visual or proprioceptive coordinates (i.e., a state of the same kind as sensory input states), is triggered both during the *mere* simulation of action *and* during normal perception (where there is reafferent input). I have already argued that, as an account of the mechanisms underlying perception, this sort of ‘comparator’ theory posits more than is needed to explain the observed attenuations of self-initiated sensory changes.¹⁵⁶ So, I disagree with simulation theorists such as Proust and Jeannerod and Pacherie, who hold that ‘internal’ or ‘stored’ reafferences are triggered ‘whether or not’ planned action comes to fruition, since such states may not be needed for normal perception (unless these ‘internal reafferences’ are understood as being the kind of filter signals I described above).

Yet, when it comes to thoughts ordinarily marked ‘imagery’ or ‘imagination’ (where imagination essentially involves sensory imagery), the idea of quasi-perceptual ‘internal reafferences’ may well be relevant. In imaging, we are able to simulate acts of perception through the triggering of ‘imagistic’ representations—exploiting what at least seem like representations of the same basic kind as those activated in perception, and which carry the same kind of information about the external world (even if, as I maintain, these kinds of representations are *not* deployed in normal perception as predictions that are compared with input). Indeed, a key difference between imagery and perception may

¹⁵⁶ It should be noted that the simulation account cannot explain how perceptual learning occurs any better than the ‘filtering’ approach I have suggested, since, on the simulation view, errors in goal-directed action are only correctable once there is a stored representation of what the correct action *would have been*. The crucial learning question concerns how this ‘correct’ representation gets to be available to the cognitive system in the first place.

be that only the former involves a willful generation of modality-specific, quasi-perceptual representations that serve as input into further cognitive processing.

The idea I shall now pursue is that inner speech is a kind of imagination or *mere* simulation of action (in this case, simulation of speaking), and, as such, can be seen as involving internally generated imagery in need of proper filtering or attenuation. But why think that imagery (and inner speech in particular) involves generating signals that serve as *input* to further processing—input that is normally attenuated or filtered? This important question can be broken into two: first, why think that imagery is in some way separate from normal thought in general, such that it could serve as input to later cognitive processing; second, why would this internally generated input normally be filtered or attenuated?

Fodor (1975, p. 175-195) provides the beginnings of an answer to the first question. Following Wittgenstein (1953, esp. *ff.* 137, fn. 2(b)), Fodor argues that images are too ambiguous to ground the intentional content we pre-reflectively take them to have. That is, what we might intuitively think of as an image representing John standing could equally well represent that John is tall, or that John has two legs, or that John is skinny, and so on, for a variety of propositions.¹⁵⁷ Fodor concludes from this that, while images are realized in a different ('iconic') representational format than 'discursive' thought, they must nevertheless have a discursive description applied to them so that they can be properly disambiguated, and then appropriately used by a cognitive system. I argued in chapter two that the kind of ambiguity Fodor points to can be explained without

¹⁵⁷ The same indeterminacy is present if we take images to represent objects instead of states of affairs (i.e., to be the bearers of *reference*, but not truth). So long as images represent in virtue of resembling their referent, an image of John is equally an image of his twin, and so on. See Fodor's example involving 'a pinwheel sort of thing' (1975, p. 182-183).

concluding that imagery occurs in a different format. Instead, I posited that mental images can be understood as existential contents (e.g., there is *a* large white arch) that are then conjoined to causal/indexical conditions (e.g. and the Arc de Triomphe looks this way) that achieve the relevant disambiguations. For present purposes, what matters is that either picture allows us to see why “imagistic” representations (i.e. mental images) might themselves serve as inputs (in need of disambiguation) for later cognitive processing.

Of course, the representational format and informational richness of visual imagery is a matter of long dispute (see, e.g., Tye (1991), Pylyshyn (2002), Slotnick *et al.* (2005)); less commonly discussed are the perhaps more developed (and certainly more commonly exploited) resources of *auditory verbal imagery* involved in the inner speech of most humans. The imagery debate aside, it should be uncontroversial that verbal imagery is tied to sensation in roughly the same way as visual imagery, whatever the ultimate nature of this link may be. Yet it is easy to overlook an important complication in the case of inner speech. We hear ourselves whenever we speak, so simulating speaking involves simulating *hearing* just as much as it involves generating thoughts and simulating various speech movements. Just try to engage in inner speech that you cannot “hear”—i.e., act out a situation *in inner speech* where you are speaking (reading a poem to a friend, say) but cannot hear yourself. I am not able to do it (I start to form visual images of my mouth moving instead). For me, it is as though there can be no inner speech that is not a simultaneous simulation of hearing myself speak; at very least, it is safe to assume that inner speech is almost always partly a simulation of auditory perception.

This link to sensation suggests that if a modular theory of mind (Fodor, 1983) is generally correct, then inner speech involves interaction between central processing and an auditory perception module (and visualization involves cooperation between central processing and a visual perception module). At the end of chapter two, I suggested that the IMG operator in front of the various imaginative contents could help to indicate that imagination involves this kind of interaction between a module and central processing. Here again—for both inner speech and visualization—there is reason to think that imagination involves internally generated inputs (i.e. internal reafferences) that could, in normal cases, be filtered or attenuated. And, as was the case with visual imagery, the signal returning from auditory perception modules may require further interpretation or disambiguation as well (am I hearing myself say thus and such, or hearing someone else who sounds like me?).

Why think that this ‘input’ to central processing would normally be *filtered*? This is the second question above. An answer can be found in two places: research on the developmental roots of inner speech, and comparative neurological data during imagery and action/perception tasks, both of which suggest that (mere) imagination or simulation shares underlying mechanisms with perception and action. Following Jones and Fernyhough (2006)—themselves drawing on the empirical work of developmental psychologist Vygotsky (1934/1987)—we can see the very ability to engage in inner speech as grounded in the ability to simulate acts of auditory perception and speech ‘offline’. On this view, we *first* learn to grasp linguistic meaning through hearing and watching others interact, then acquire the ability to speak. Only later are we able to engage in inner speech through simulating these learned abilities, exploiting a store of

auditory verbal imagery in the process. If one's own normal speech is automatically accompanied by filter/efference-copy commands that attenuate one's sensitivity to the sound of one's own voice (as was found by Blakemore *et al.* (2003)), and if inner speech, due to its developmental roots, must draw on the same abilities and mechanisms as 'out loud' speech, then it is plausible to suppose that inner speech is normally accompanied by filter/efference-copy signals as well.

Measurements of brain activity during perceptual and imagery tasks also bear out the hypothesis that a common set of mechanisms underlies both overt (e.g., speech, perception) and covert (e.g., inner speech, visual imagery) actions. Jeannerod and Pacherie (2004, p. 130) discuss results from neuro-imaging studies showing activity in a common neural network activated during both (mere) motor imagery and action preparation. Currie and Ravenscroft (1997) cite a wealth of similar data: for instance, by measuring regional cerebral blood flow, Roland and Friberg (1985) and Goldenberg *et al.* (1989) showed that common cortical regions were active both during visual perception and visual imagery; measuring alpha rhythms, Davidson and Schwartz (1977) showed that activity increased in the visual cortex during visual imagery. More recently, Slotnick *et al.* (2005) found similar visual imagery and perception tasks to evoke cortical activity with the same precise visual field topography ('retinotopy'), and to cause common effects in motion processing region MT+. Kosslyn and Thomson (2003) present a meta-analysis of a variety of neuroimaging studies with similar findings.

The other main source of neurological evidence that imagery and perception involve common mechanisms comes from patients with perceptual or motor deficits that are mirrored in their abilities to use imagery. Sirigu *et al.* (1995) found that deficits in a

patient's ability to move her fingers were attended by similar deficits in her ability to imagine moving her fingers. Corresponding data has been found regarding visual imagery and vision (Brooks, 1967): patients who, due to brain lesions, neglect a certain region of the visual field will neglect that same region in visual imagination. Another suggestive link stems from the well-known work of Kosslyn *et al.* (1978), who found that when subjects are asked to imagine making a movement, the time taken closely mirrors that taken in reality (and, similarly, that difficult motor-actions take longer to perform in imagination just as they do in reality), again implicating a common set of mechanisms. Notably, for schizophrenic patients experiencing inserted thoughts and voices, imagined movements tend to take the same time whatever the level of difficulty (Danckert *et al.*, 2002, p. 605-609).

Turning finally to the auditory imagery of inner speech implicated in auditory hallucinations and inserted thoughts, Stephane *et al.* (2001, p.75) present evidence that 'inner and external speech have the same developmental origin and both have perceptive and motor components.' To cite one striking example from their survey, Dell (1978) found that 'tongue-twisters', such as the phrase 'unique New York', cause the same sort of errors when 'uttered' only in inner speech as they do when said aloud. Looking at brain activity, Ford and Mathalon (2004) found a dampening of auditory cortex responsivity during both talking and inner speech in normal subjects, but found no such dampening of either activity in schizophrenic patients prone to auditory hallucinations.

The neurological and developmental evidence linking 'out-loud' speech to inner speech, combined with the view of inner speech as a kind of imagining involving the interaction between sensory modules and central processing, suggests the following

picture. Inner speech is accompanied by efference copy signals just as normal speech is (because the former draws on the same underlying neural mechanisms and abilities as the latter); and to generate inner speech is to generate a kind of *imagery*, which requires interaction among multiple cognitive modules. For thought to become inner speech, central processing must draw upon the resources an auditory memory module. It is the signal returning from this module that serves, in lieu of perceptual feedback, as the relevant 'reafferent' signal to be attenuated by the efference-copy triggered filter command. Defects in the filtering of this signal may lead to the abnormal phenomenology associated with thought insertion.

In closing this section, it should be noted that some have challenged the proposed link between inner speech and the 'voices' heard by schizophrenic patients. In a series of elegant experiments, Evans *et al.* (2000) showed schizophrenic patients prone to AVHs not to be impaired, relative to patients without AVHs, in various tasks involving the interaction of the 'inner voice' and 'inner ear'. In one experiment, patients prone to AVHs proved equally skilled to non-hallucinators at determining which pairs of written words were homophonic, which rhymed, and which 'sounded totally different' (they were asked to accomplish the task by silently 'pronouncing' the words in inner speech) (2000, p. 144). Evans *et al.* concluded that inner speech and auditory verbal hallucinations are not connected 'in a simplistic or direct way' (2000, p. 137). Yet, since it is well known that AVHs are significantly decreased by tasks that draw on or 'preoccupy' the normal mechanisms related to inner speech (Bick and Kinsbourne, 1987; Nelson *et al.*, 1991; Margo *et al.*, 1981), an alternative interpretation of their data is that the presented tasks sufficiently taxed the mechanisms underlying inner speech as to

suppress normally defective corollary discharge responses. Evans *et al.* discuss this possibility themselves, noting that AVHs ‘appear to be spontaneous and unbidden,’ and, therefore, ‘it is possible that our tasks did not tap into this potentially crucial feature of AVHs’ (2000, p. 146). Thus, their findings do not close the door on inner speech-based theories of AVHs and inserted thoughts, even if they indicate that a mere appeal to ‘deficient inner speech’ is not explanatorily sufficient. In particular, an explanation of the intermittent character of the posited inner speech deficits will eventually be required. This issue is discussed further in section six below.

5.2 The Filtering Hypothesis and Inner Speech – Phenomenological Consequences of Filtering Deficits

So far I have explained how sensory attenuation may occur through efference-copy generated filter signals, and have given reason to think that the same principles can be extended to covert actions such as inner speech. It remains to be explained how a lack of filtering in the case of inner speech could result specifically in a phenomenology that leads one to think that one’s own thoughts are someone else’s. That is the project of this section.

A lack of filtering could give the phenomenology of inner speech an element of extraneity in a variety of ways. Just as with vision, there may be distinctive *structures* or *patterns of change* that correlate with ‘filtered’ inner speech (and with hearing oneself speak) that are different from those associated with hearing others speak (the patterns being reversed in the case of inserted thoughts). Think again of the eye-nudging case: it is not that one’s visual experience while poking one’s eye is more or less vivid or rich

than during normal visual exploration. The salient phenomenological difference is rather a structural difference in the nature of the respective signals across time—specifically, the difference consists in how similar successive states of the (downstream, conscious) visual system are to each other in the nudging and non-nudging scenarios. When efference copies are triggered during normal visual exploration, there will at times be considerable gaps (in terms of the type of information nested) between successive visual states, due to the filtering of the signal as the organism saccades (resulting in a ‘stable’ visual field). Conversely, the ‘blurring’ that occurs when efference copies are suppressed (in the eye-nudging case) results from the relative similarity of information nested in each successive state during active perception—put simply, less of the signal across time is ‘thrown out’.

Similar structural differences may exist between filtered and unfiltered inner speech. One possibility is that filter signals normally prevent inner speech from being fully enunciated, both in order to speed its progression and to decrease the cognitive resources needed for its processing. When the filter signals are absent (due to an efference copy defect) one’s inner speech may be fully enunciated, slowing its progression and giving it an alien quality. By subtracting unnecessary information from the signal, filtering would enable normal inner speech to be easily suppressible in the event of unexpected perceptual input from outside (after all, even if central processing needs to draw on stored auditory information or abilities to get inner speech going, this does not mean that it needs everything that comes back from these modules). If filtering is lacking, inner speech may not be automatically ‘overlooked’ when other inputs arise,

and may itself seem to be one of those externally-derived inputs, not a self-generated signal.¹⁵⁸

These suggestions for the precise phenomenological consequences of failure to filter inner speech are only schematic. They serve principally as pointers toward the correct approach to understanding the phenomenology of agency. We face a major obstacle in explaining what it is like to have this phenomenology in any fine-grained way because most of us do not know what it is like to have such experiences. When it comes to sympathetically imagining inserted thoughts, we are in the same position that someone who has never rubbed his eyes (or experienced an illusion) would be in were he asked to imagine his visual field ‘wobbling’. It seems it would be quite difficult for such a person to understand what was meant by a ‘visual field’ that can ‘move’ independent of objects moving in the environment.

This is overlooked in most discussions of the phenomenology of thought-agency; implicit in such discussions is the assumption that the *lack* of this phenomenology is something most people are familiar with or can easily imagine. The best we can do, I suggest, is point to structural and qualitative features this phenomenology may have, based on cases with which we are familiar. These features are still explanatory, however, since certain structural differences in phenomenology may plausibly be associated with the detection of changes in one’s environment, just as the visual field wobble is a visual

¹⁵⁸ The ‘hyperreflexivity’ posited by Sass and Parnas (2003) and Gallagher (2004)—whereby the apparent extraneity of inserted thoughts arises out of schizophrenic patients becoming over-aware of their own thinking—may actually be a result of a failure of filter commands to meet and attenuate the signals returning from centers involved in generating auditory imagery. If that is right, then the putative hyperreflexivity consists not in an introspective faculty ‘looking too hard,’ but from the internally generated signal’s unusual ‘strength’ (i.e., the increased amount of cognitive resources required for its processing).

phenomenology naturally associated with the visual perception of the entire room shifting to the side.

Notice that this account of the phenomenology of thought *agency*—and of its possible disorders—places that phenomenology within the auditory sense modality, as being embedded in certain filtered structures and patterns of auditory verbal imagery. Similarly, the phenomenology of agency can also at times be *visual* in nature, such as when intended eye movements result in efference copy commands that lead to a stabilized visual field; the filtering out of information received between the beginning and end of a saccade results in a kind of agency phenomenology that is correlated with willful acts of visual exploration. This kind of agency-phenomenology is not contained within imagistic ‘snapshots’, but rather arises out of distinctive patterns of information across time. If the input information between the beginning and end of a saccade is not filtered, the downstream pattern of information—across time—will be different; it will constitute a phenomenology associated with changes in one’s environment.

Other theorists (Proust, 2006; Jones and Fernyhough, 2006) apply a similar efference-copy/reafference paradigm to explain how inner speech is mistaken as being externally generated, yet hold that an ‘emotion of self-authorship’ is the relevant aspect of phenomenology that is altered or missing because of efference copy defects. This is a crucial misstep, since it requires that we have some idea of what this emotion is in the normal case. But, as noted at the outset, it is far from clear what this emotion might be, and if it truly attends most normal thought. To suggest that it is normally unnoticed or taken for granted is to say it is not part of non-pathological *experience*, hence not an element of phenomenology after all.

The appeal to an ‘emotion of self-authorship’ is motivated by comparator models that place the relevant ‘mismatch’ below the level of consciousness. On such accounts, it is only an unusual emotion (or lack thereof) that bubbles to the surface in schizophrenic experience, the patient’s sensory phenomenology being otherwise normal. The account I have given does not appeal to peculiar emotions, but rather to structures of changes in downstream sensory information—changes that are themselves present to consciousness as a part of what Gallagher (2004) calls ‘first order phenomenology,’ and which may plausibly lead to reports of thought insertion. This way, a *distinct* form of unusual phenomenology is described—one which sets schizophrenic cases apart from the normal lack of agency attending things such as melodies stuck in one’s head. The phenomenology of agency that concerns us is therefore one that is embedded in all first order sensory and proprioceptive phenomenology as diachronic, action-sensitive patterns of information; it does not stand apart from them as an inscrutable emotion.

5.3. Fractured Phenomenologies – Unified Mechanisms for Inserted Thoughts and Auditory Verbal Hallucinations

Earlier I spoke of the way in which one’s overall phenomenology becomes ‘fractured’ in the eye nudge case. The fracturing is due to a conflict between what the visual system is registering and what other sensory and proprioceptive inputs indicate. It is worth reflecting on the kind of fracturing that (if the present hypothesis is correct) takes place in the case of inserted thoughts. Interestingly, there does not seem to be a *direct* auditory equivalent of a visual field; that is, there is no ‘auditory field’ that seems to wobble when one has an auditory experience as of someone speaking, while the inputs of other sensory

channels indicate there is no such speech occurring. We can suppose, however, that inner speech and the perception of one's own out-loud speech are filtered in various (attenuating) ways that set them apart, phenomenologically, from the perception of the speech of others (this may be part of the reason our voices sound different to us when played back from a recording). When this filtering is lacking, inner speech may be characterized by the kind of unfiltered phenomenology normally attending the perception of another's speech.

Neurological data backs up these speculations. Ford and Mathalon (2004) found that, in schizophrenic patients prone to AVHs, responsivity of auditory cortex during inner speech is similar to that during normal auditory perception (in normals inner speech is accompanied by a dampening of responsivity in auditory cortex). Dierks *et al.* (1999) found that when auditory hallucinations occur (as indicated by the patient) brain metabolism increases in the primary auditory cortex on the left side, indicating that the nervous system in these patients behaves as it would during normal speech perception (they also note that inner speech is normally attended by a decrease in primary auditory cortex metabolism). And Daprati *et al.* (1997, p. 73) found that 'during verbal hallucinations in schizophrenic patients...the sensory areas for language remain active, which suggests that the [normal] cancellation process does not operate.' Assuming that, when this lack of cancellation occurs, other sensory and proprioceptive inputs conflict with the notion that someone nearby is speaking, a fracturing of one's overall phenomenology results.

Yet there is also a *second* fracturing that can occur in the phenomenology of inner speech. The second fracturing has to do with the comparatively impoverished sensory

character attending unfiltered inner speech. Because the sensory character of inner speech is not nearly as rich as that of actual speech perception, an act of inner speech is usually phenomenologically distinguishable *as inner speech* partly by this paucity of sensory character. Yet, when this lack of sensory paucity is coupled with a lack of normal filtering, the inner speech at once possesses an element of innerness (the lack of sensory character) and outerness (the lack of filtering). This constitutes a second form of fracturing.

These two forms of fracturing would place unfiltered inner speech in a very peculiar phenomenological category. Experiencing such inner speech would be somewhat like hearing another speak, and somewhat like engaging in inner speech, but not exactly like either. How would one describe such an experience? Perhaps as an ‘inserted thought’ (due to its lack of sensory character), yet it seems equally likely that one might describe it as an episode of seeming to hear someone speak (due to its lack of filtering). My suggestion, therefore, is that AVHs and inserted thoughts may at bottom be the same phenomenon reported in different ways, precisely because their phenomenology straddles the fence between speech perception and inner speech.

The usual approach to understanding these phenomena is to assume we have a reasonable idea of what AVHs are like (they are just like hearing someone speak) which leads to a puzzle over what might account for reports of inserted thoughts. I suggest instead that we question whether most AVHs really are just like hearing someone else speak, and instead assimilate them to the fractured phenomenological realm just described. Given the co-presence of these putatively distinct symptoms in many patients (Nayani and David, 1996), a unified account of their source is desirable. Crucially, this is

a unification that takes seriously the reports of patients who claim their inserted thoughts are *not* simply like cases of hearing someone speak.

6. Loose ends: Intermittency and Content

There are two important characteristics of inserted thoughts that have not been explained by the present account. The first is their intermittency. Presumably, not *all* of the inner speech of schizophrenic patients is improperly filtered and lacking in the usual phenomenology of agency. But if the inserted thoughts are caused by a deficit in filtering, why wouldn't this deficit afflict all of patients' inner speech? The short answer is that these filtering errors must themselves be intermittent. Admittedly, this leaves the intermittent character of inserted thoughts unexplained. But this should not worry us, as it is a normal feature of explanations of intermittent phenomena that they first locate the causal factors, and then *separately* give an account of why these causal factors are only intermittently active. For instance, my intermittently runny nose may be explained by the presence of high levels of histamines in my bloodstream. Once we know that histamines are responsible for my runny nose, we can then ask why these histamines are only occasionally present in the high degree needed to induce a runny nose. A separate account is then given explaining how contact with intermittently present allergens leads to the release of antibodies which in turn cause an increase in histamine levels.

Note that a key part of arriving at the explanation of intermittency in this case is that the biological basis of histamines is properly understood, for this is what allows the histamine explanation to be linked to a further explanation involving antibodies and allergens. Similarly, we should expect that, once the biological or neurological basis of

the posited filtering mechanisms is located, we will be in a much better position to explain why these mechanisms are only intermittently unreliable.

A second worry may be that the present account does nothing to explain the particular content of inserted thoughts—i.e., why do such thoughts have one subject-matter and not another? If it were the case that all or most inserted thoughts had a particular content—were personally abusive, say—there would be good reason to suppose that this content plays a causal role in generating the relevant abnormal phenomenology. As reports of the contents of large numbers of actual inserted thoughts (as opposed to auditory verbal hallucinations) are hard to come by, it is difficult to assess this hypothesis. We do know, however, that AVHs often have a positive or comforting content for their bearers (Bleuler 1911/1950). When asked whether they ever experienced ‘pleasant’ voices, forty-eight of 100 schizophrenic patients surveyed by Nayani and David indicated that they did (1996, p. 182). Nayani and David also found that, while the development of coping strategies by patients was associated with the presence of pleasant voices, this change in content was not sufficient to stop the hallucinations (1996, p. 186). Thus, any account which tries to explain inserted thoughts in terms of anxiety-ridden or displeasing content will have to give a separate account of the nature of AVHs. A virtue of the filtering account is that it can treat inserted thoughts and AVHs as a single experiential phenomenon. The filtering hypothesis’s approach to the question of content will be to predict no special link between thought content and a lack of a sense of agency.

By avoiding an appeal to content as a causal factor, the filtering approach remains broadly applicable in a way that allows it to explain other schizophrenic symptoms as

well, such as delusions of control (where the subject feels her bodily movements not to be under her own control), and the well-documented failures to attenuate the consequences of self-initiated changes in a variety of sense modalities (Shergill *et al.*, 2005). An account of inserted thoughts that focuses on their content as causally relevant (e.g., Vosgerau and Newen (2007, p. 38-40)) could not extend itself to these cases in any clear way. Given that inserted thoughts rarely occur in isolation, but rather are just one of several typically co-present symptoms of schizophrenia, it is highly desirable to develop an explanation applicable to more than one of these co-present symptoms.

7. *Virtual Reality, 'Presence', and Belief Fixation*

With this proposal in hand, how close are we to a solution to the puzzle of extraneity? One might worry that, despite having a firmer grip on the possible phenomenology and etiology of thought insertion and auditory hallucinations, it is still hard to see what would cause schizophrenic patients to continually form such delusional *beliefs*. If their general ability to reason is not impaired, why do patients so strongly resist concluding that they are simply perceptually unreliable, persisting instead in attributing the agency of their thoughts and 'voices' to someone else? After all, it seems one can imagine having experiences like those of schizophrenic patients without thereby coming to believe that someone else is speaking or inserting thoughts into one's mind. Would we not, in such circumstances, simply conclude that we'd become perceptually unreliable?

Part of what accounts for such puzzlement may be an unrealistic view of the relation between perceptual phenomenology and belief. When we imagine it possible to

have the same perceptual phenomenology as schizophrenic patients while lacking their beliefs, we tacitly assume that perceptual phenomenology is cleanly separable from belief. However, the fact that sensory modules process information prior to its involvement in belief fixation does not entail that *perceptual phenomenology* (a conscious, and relatively late state in the information-processing chain) can itself completely float free of belief. It is not as though we normally *choose* what to believe based on perceptual phenomenology; perceptual phenomenology just is a kind of belief—albeit, one that can be trumped in special circumstances by extenuating background beliefs.

If perceptually-formed beliefs are in large measure formed pre-consciously, how does belief relate to early perceptual input? This is a central question of researchers developing virtual environments (VEs). Since present technologies involved in the creation of VEs cannot approach the level of informational detail and computational speed needed to create a full-on replication of the sensory consequences of perceiving a genuine environment, a crucial task is to discover the minimal perceptual cues needed in order to make a VE convincingly ‘present’ to a perceiver. ‘Presence’ is a technical term in the VE literature used to refer to a VE’s ability to instill in the perceiver a sense of actually ‘being there’ in the VE. Presence is measured in a variety of ways, the two most common being questionnaires (given to subjects after the fact), and measurements of physiological markers such as EKG and galvanic skin response during the exploration of a VE (Sanchez-Vives and Slater, 2005).

Most VEs require subjects to wear a head-mounted display (HMD) which funnels computer-generated visual information to the left and right eyes via small screens (each

eye receives slightly different signals in order to induce the proper three-dimensional effect). Crucial for presence in a VE is that the rate at which graphics are refreshed is at least 15Hz, and that these changes occur in accordance with the subject's movements (Sanchez-Vives and Slater, 2005; Barfield and Hendrix, 1995). For instance, in a task where subjects were asked to retrieve an object from a (virtual) room with a deep 'pit' at its middle (leaving only a foot or so of clearance on any side of it for one to make one's way around to the goal object), heart rates were significantly increased when the frame-refresh-rate was over 15Hz, yet did not significantly change when frame rates were lower (here heart rate is being used to judge presence) (Meehan *et al.*, 2002).

Interestingly, the visual realism of a virtual environment has been shown to be a negligible factor for the creation of presence (Usoh *et al.*, 2000; Zimmons and Panter, 2003; Sanchez-Vives and Slater, 2005). Crudely rendered objects, persons, and rooms can achieve high ratings on all tests of presence, so long as the screen shots of such environments are updated at the required rate, and are done so pursuant to the actual head and body movements of the subject. Two other factors shown to significantly increase presence are a representation of the body of the subject herself (even if crudely or oddly rendered), and the ability of a subject to actually walk about (as opposed to only looking) while engaging with the VE (Sanchez-Vives and Slater, 2005).

These findings have at least two important consequences for thinking about the puzzle of extraneity. The first is that our pre-theoretical intuitions about the factors in perception that lead to the fixation of belief may not be reliable—a richly detailed environment may be relatively unimportant for the formation of belief-like states, so long as certain patterns of information are received at the proper rate by the early perceptual

system. Second, this work suggests a strong link between the sense of *extraneity* one has regarding a VE and the receiving of certain action-sensitive patterns of sensory stimulation—patterns that may arise partly out of specific filtering algorithms being applied (or not). Even though the subjects ‘know better’ than to become nervous in front of an unhappy-looking audience¹⁵⁹, or to fear the virtual ‘pit’ in the center of a room, a degree of belief that these scenarios are real nevertheless attaches, resulting in increased heart rates and impeded performance. As the technologies underlying VEs become more sophisticated, opportunities will arise to further test the filtering hypothesis, as perceptual feedback from the VE could conceivably be manipulated in ways to duplicate in normal subjects the effects postulated to occur in patients with efference copy deficits.

More importantly, VEs may eventually offer a new form of therapy for those suffering from the hallucinations and delusions characteristic of schizophrenia. VEs have already been used to help treat some phobias. Patients are submerged in virtual environments simulating their phobias to various degrees of intensity, allowing them to gradually become accustomed to facing such situations without anxiety (Rothbaum *et al.*, 1999). Specific perceptual-motor deficits—such as those attending Parkinson’s Disease—are also treated using VEs (Prothero, 1993). While the VE does not stimulate increased production of dopamine, it does allow for the discovery of new coping strategies by presenting disability-specific challenges (Wann, 1996).

¹⁵⁹ Sanchez-Vives and Slater (2005) asked their subjects to deliver a speech to two different virtual audiences. One was ‘positive’—the listeners’ eyes programmed to remain open and fixed on the subject’s, the virtual characters making nodding movements and sounds of approval throughout the talk. The other was ‘negative’—audience members avoided eye contact and assumed postural positions indicating a lack of interest. The virtual humans in the ‘audience’ were very primitively rendered, showing only basic facial expressions and assuming relatively stiff, mechanical postures. Yet, even experienced speakers among the subjects reported being flummoxed by the ‘negative’ audience, finding themselves emotionally upset and performing far worse than normal, while speakers addressing the ‘positive’ audience performed well and found their experiences gratifying.

These two kinds of therapy could be combined in the treatment of schizophrenic symptoms. The sort of ‘unfiltered’ sensory feedback thought to attend some actions of schizophrenic patients can be independently generated by manipulating the behavior of the VE as it is explored, allowing the patient to become more familiar with certain types of dissociations. Becoming accustomed to irregular feedback in a VE may both decrease the patient’s anxiety in the face of similar pathological experiences, and may allow for the discovery of novel coping strategies—strategies that can only be discovered in an atmosphere lacking the anxiety typically surrounding pathological experience. Even if a VE cannot duplicate the phenomenology of unfiltered inner speech (since the only inputs are normal sense organs), growing accustomed to unfiltered visual and auditory input (in situations where efference copies should lead to filtering) may still allow for the development of coping strategies relevant to dealing with unfiltered ‘internal’ stimuli.

8. Conclusion

There is no short or obvious path toward answering the puzzle of extraneity. The overall strategy adopted herein has been to take the first person reports of schizophrenic patients seriously, and to search for the simplest way of understanding them in terms of the mechanisms underlying normal perception and thought. I have given reasons to think that thought insertion and auditory hallucinations may result from single general deficiency—a lack of the proper ‘filtering’ of inner speech—and, therefore, that these two cardinal symptoms of schizophrenia may themselves best be thought of as marking distinctions in degree of a single underlying cognitive disorder. Importantly, the mode of explanation employed—invoking efference copy-driven sensory cancellation—is one that

can be extended to explain schizophrenic hallucinations and delusions of control within other proprioceptive and sense modalities as well.

On the view I have proposed, we should locate the phenomenology of agency *within* the traditional sensory and proprioceptive modalities, as diachronic patterns of high-level processing, not outside of them as an extraneous emotion or ‘feeling of effort’. Such patterns *just are* the extraneity that attends schizophrenic experience and which, in non-pathological cases, allow one to distinguish externally from internally caused changes in sensation. Understanding why these patterns result in the specific delusions and hallucinations of schizophrenia, I have argued, requires looking at the way overall phenomenologies are shaped by the interaction among multi-modal sensory and proprioceptive inputs. Doing so allows one to see the fault lines along which phenomenologies may fracture.

SUMMARY AND SPECULATION

From Descartes forward, the ability to imagine or conceive of a situation has often been considered a reason to grant its possibility. Even those who have denied the ability of imagination to warrant modal judgments have still thought of it as a belief-independent faculty, something that runs cognitively *free*, unfettered by any one's actual commitments. I have argued for an opposite view; imagination, both sensory and propositional, is *constituted by* our commitments. It involves assessing the world to be various ways—most saliently, it involves assessing possibilities.

Thus, considering one's (apparent) ability to imagine this or that scenario as evidence for its possibility is akin to considering one's belief that the same is possible as evidence (or warrant) for its own truth. But beliefs rarely, if ever, justify themselves. As discussed in Chapter Three, some imaginings—some assessments of possibility—may be justified by the chain of imaginings that lead up to it, just as a belief may be warranted by the chain of inferences that lead to it. But the ability to imagine things does not entail an ability to reach *outside* of our collective commitments to a new source of information, which can then be used to justify new commitments. Imagining just is exercising our existing commitments in order to form new ones, be they antecedently justified or not.

This basic insight was expressed somewhat differently for sensory imagination than it was for propositional imagination. In the case of sensory imagination, I argued we can easily be wrong about what we are imagining, even if we have privileged access to what we seem to be imagining (or, equivalently for our purposes, to what we believe we are imagining). In some cases of modal imagining, we will be wrong about what we are

imagining simply because a modal commitment that forms part of the content is false. Unlike the assertoric representations involved in non-imagistic propositional imaginings, sensory imaginings are not themselves beliefs. I argued (in Chapter Two) that this must be so given the dissociations revealed in the empirical literature between the ability to sensorily imagine an object and to form beliefs about the same object; someone can believe that it is possible for the Empire State Building to be painted red without being able to (modally) imagine it painted red, so (modally) imagining the Empire State Building painted red cannot be the same thing as (occurrently) believing it could be red.

In the case of propositional imagination, the key mental states involved are beliefs in conditionals. To imagine that cancer is cured is to infer a variety of beliefs of the form, “If cancer were cured, then X,” from one’s standing beliefs. The freedom of imagination consists in our ability to choose the subject matter of this reasoning, including reasoning about what would happen in scenarios that are themselves unlikely.

In closing, I will say a little more about two issues relating to sensory imagination that were quickly passed over in earlier chapters. The first concerns the representational “format” of imagery. In Chapter Two I remained neutral between different (“descriptivist” and “pictorialist”) theories of imagery format, focusing instead on the matter of *what* images represent. Yet Chapter Five introduces ideas that may ultimately weigh against descriptivist approaches. I argued that the fast attenuation of self-generated changes in perceptual and imagistic inputs occurs through a simple filtering process. For instance, the same filtering algorithm can be applied to any

visual input resulting from an instance of looking to the right, since the same pattern of attenuation of the signal across time will suffice for the stabilization of the visual field.

If mental images are like sentences, it is hard to see how they could be “filtered” in the way proposed. Filtering out words here and there from such a “sentence” would usually result in something ungrammatical, which would remove the rationale for thinking of the representation as sentence-like. Nor is there any reason to think that the same filter could apply to any sentence generated by, say, a case of looking to the right. To arbitrarily delete a term of a sentence would remove from the representation an object (or property) in its entirety, rather than simply removing excess information relevant to representing the object or property (as a filter plausibly would).

This does not, however, mean that pictorialism is likely true. The problem with pictorialism lies in the vacuity of comparing mental images to pictures (or photographs). Pictures and photographs are colored and are ordinary (more or less) two-dimensional objects. Mental images are not themselves colored (they merely represent things as being colored), nor can they be said to *look* like pictures, since we do not see mental images any more than we see perceptual experiences. Recent work in cognitive psychology has shown that at least some mental images may result from retinotopically organized activation in early visual centers of the brain, providing some substantiation to the idea that, like pictures, they have a (roughly) two-dimensional shape. While an interesting result, the question is whether anything is *added* to our understanding of mental images by pointing out that pictures also have a two dimensional shape. For instance, the pages of books have a (roughly) two-dimensional shape, but, obviously, this in itself does not warrant a “pictorialist” account of how the sentences in books refer.

What really lurks behind the comparison of mental images to pictures is the old thought that both achieve reference in virtue of resembling their referents. But as Goodman (1976) pointed out, resemblance is a symmetric relation, while reference is asymmetric: if A resembles B, then B resembles A; but A referring to B does *not* entail that B refers to A. Clearly any story concerning *how* images manage to represent will have to go beyond resemblance claims. And, even given this further story about why such resemblances counts as cases of representation, the resemblances that do hold are too coarse-grained to explain the rich representational role that images play in memory and planning. After all, the same pattern of activation on the retina (itself transferred to retinotopically-organized areas of visual cortex) can result from perceiving—and equally well resemble—three-dimensional objects of a wide variety of shapes and orientations. Thus, even an object's *shape* could not be represented solely in virtue of a resemblance between the (neurally-instantiated) mental image and the object itself—why, then, think that resemblance is in any way relevant to answering the question of how (and what) an image represents?

More confusion than elucidation likely results from any comparison of mental images to pictures. A third way must be developed, one which starts with the idea that mental images are not importantly sentence *or* picture-like.

To close this dissertation, I will expand a bit on some thoughts offered at the end of Chapter Three. There I speculated that much of the bewilderment over the mind/brain relation results from the improper use of imagery to imagine mental states. Image-involving thought cannot provide reliable information concerning what mental states

themselves are like, I argued; for images, like percepts, represent worldly objects and properties—they are what we use to think about worldly objects and properties, and they give us reliable information about such objects and properties. But having images and percepts does not help us understand, or think about, *their own* nature.

Put more provocatively, perceptual experiences—and mental states in general—*are not imaginable*. That is, they are not *sensorily* imaginable; and yet sensory imagination seems to underlie so much theorizing about the mind/brain relation. The many recent attempts to carve out a special category of “phenomenal concepts” (e.g., Papineau (2002)) that are constitutively connected to imagery, and which somehow or other account for the appearance of a mind/brain gap, work too hard. The confusion is simple: we use images while thinking about other people’s experiences not because their experiences possess the luminous properties represented by our images, but because we want to know what they are thinking about. So, if I see you peering over a wall, I may form an image of various objects on the other side, and, in an offhand way, attribute this image to you. What is correct in this process is my attributing to you an awareness of various properties and objects on the other side of the wall that are also represented by my image. What is wrong is supposing that those properties themselves—or even special mental analogs of them—are somehow instantiated in your mind.

The standard objection to such deflationary remarks is that the possibility of spectrum inversion—of a person having the same kind of experience when seeing red that someone else has seeing green, etc.—shows that there really are mental qualities *akin* to colors that are not reducible to mere representations of physical colors. But my point aims to defuse this intuition before it is lit. The question is why we think such inversion

is possible in the first place. I say it is due to an improper use of imagery—a sloppy ascription of properties *sort of like* the properties actually represented by the image to the mental state one is thinking about while having the image. This is what accounts for the conscious mind’s seeming rather well-lit, in possession of its own “mental colors”.

The question, then, is why we do not usually think about non-human representation-formers in the same way (to the extent that we don’t). Consider Blink the Robot, a simple silicon soul who spends his days scanning the environment with a video camera and, with the help of commercially-available feature-recognition software, reports on nearby objects and colors. Common wisdom has it that, for Blink, *all is dark inside*. But is this fair to Blink?

Put Blink on the beach, sitting cross-legged on a blanket next to you and a loved one. It has been a long day of fun, catching waves, playing paddle-ball, and reporting on nearby objects and colors. The air is cooling; a magnificent sunset unfolds. In the distance a jet crosses the sky, leaving behind it a bright trail of orange smoke, as though pulling back the peel of a ripe nectarine. Richard Branson appears above you in a giant hot air balloon, then floats away again, continuing on in his remarkable quest.

You look at Blink and think about all he is representing: the incredible colors, the departing hot air balloon, the rolling tide. You look to your loved one, who is similarly taken with the magical scene. Considering her experience, you think: all of *that*—the sea, the sky, the Virgin mogul—all of that is in *there*, in her mind, a free-wheeling movie on a frameless, transparent screen...just as it is in mine. The richness of human experience!

And then you look again to Blink who, truth be told, appears less rapt. “Orange!” cries Blink. “Jet!” he exclaims. “Hot air balloon in the way,” he drily notes. Having still

the rich image of all that is before you in mind, lingering from your meditation on the visual experience of your human companion, you wonder: should I attribute *all this* to Blink? All of this richness? Or should I deny him this bountiful image since, as they say, all is dark within him?

As you think this over, Blink looks at you. Blink blinks. “If it’s dark in me,” he says, “it’s dark in you.”

REFERENCES

- Adams, E. W. (1975). *The Logic of Conditionals*. Dordrechtss, Holland: Reidel Co.
- Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45(7), 1363-1377.
- Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-related changes in the episodic simulation of future events. *Psychological Science*, 19, 33-41.
- Arp, R. (2005). Scenario Visualization: One Explanation of Creative Problem Solving. *Journal of Consciousness Studies*, 12, 31-60.
- Arp, R. (2006). The Environments of Our Hominin Ancestors, Tool Usage, and Scenario Visualization. *Biology and Philosophy*, 1, 95-117.
- Arp, R. (2008). *Scenario Visualization*. Cambridge, MA: MIT Press.
- Barfield, W. and Hendrix, C. (1995). The effect of update rate on the sense of presence within virtual environments. *Virtual Reality: The Journal of the Virtual Reality Society*, 1, 3-16.
- Baron-Cohen, S. (1989). The autistic child's theory of mind: a case of specific developmental delay. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 30(2), 285-297.
- Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press.
- Baron-Cohen, S., Leslie, A., & Frith, U. (1986). Mechanical, behavioral and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology*, 4, 113-125.

- Baron-Cohen, S., Wheelwright, S., Stone, V., & Rutherford, M. (1999). A Mathematician, a Physicist and a Computer Scientist with Asperger's Syndrome: Performance on Folk Psychology and Folk Physics Tests. *Neurocase*, 5, 475-483.
- Bartolomeo, P., Bachoud-Levi, A. C., & Denes, G. (1997). Preserved imagery for colours in a patient with cerebral achromatopsia. *Cortex*, 33(2), 369-378.
- Bealer, G. (1994). Mental Properties. *The Journal of Philosophy*, 91(4), 185-208.
- Behrmann, M., Moscovitch, M., & Winocur, G. (1994). Intact visual imagery and impaired visual perception in a patient with visual agnosia. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 1068-1087.
- Bennett, J. (1995). Classifying Conditionals: The Traditional Way is Right. *Mind*, 104(414), 331-354.
- Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Clarendon Press.
- Bermudez, J. L. (2008). Noconceptual Mental Content. *The Stanford Encyclopedia of Philosophy*. Retrieved from:
<http://plato.stanford.edu/archives/spr2008/entries/content-nonconceptual/>
- Bick, P.A. and Kinsbourne, M. (1987). Auditory hallucinations and subvocal speech in schizophrenic patients. *American Journal of Psychiatry*, 144, 222-225.
- Blakemore, S.J., Wolpert, D.M., and Frith, C.D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, 1, 635-640.
- Blakemore, S.J., Wolpert, D. and Frith, C.D. (2002). Abnormalities in the awareness of action. *TRENDS in Cognitive Science*, 6, 237-241.
- Blakemore, S.J., Oakley, D.A. and Frith, C.D. (2003). Delusions of alien control in the normal brain. *Neuropsychologia*, 41, 1058-1067.

- Blakemore, S.J. (2005). Recognizing the sensory consequences of one's own actions and delusions of control. In T. Feinberg and J. Keenan (eds.), *The Lost Self*. Oxford: Oxford University Press.
- Bleuler, E. (1911/1950). *Dementia Praecox or the Group of Schizophrenias*, trans. J. Zinkin. New York: International Universities Press.
- Block, N. J. (1981). *Imagery*. Cambridge, Mass.: MIT Press.
- Boyd, R. (1992). Materialism with Reductionism: What Physicalism Does Not Entail. In N. J. Block (Ed.), *Readings in Philosophy of Psychology* (Vol. 1, pp. 67-106). Cambridge, MA: Harvard University Press.
- Brann, E. T. H. (1991). *The World of the Imagination: Sum and Substance*. Savage, MD: Rowman & Littlefield.
- Brindley, G. and Merton, P.A. (1960). The absence of position sense in the human eye. *Journal of Physiology (Lond.)*, 153, 127-130.
- Brooks, L. (1967). The suppression of visualization by reading. *Journal of Experimental Psychology*, 19, 289-299.
- Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *TRENDS in Cognitive Sciences*, 11(2), 49-57.
- Burge, T. (1991). Vision and Intentional Content. In E. LePore & R. VanGulick (Eds.), *John Searle and his Critics* (pp. 195-213). Oxford: Blackwell.
- Byrne, A. (2007). Possibility and Imagination. *Philosophical Perspectives*, 21(1), 125-144.
- Campbell, J. (1999). Schizophrenia, the space of reasons, and thinking as a motor process. *The Monist*, 82, 609-625.
- Campbell, J. (2002). *Reference and Consciousness*. Oxford: Oxford University Press.

- Campbell, J. (2002). The ownership of thoughts. *Philosophy, Psychiatry, and Psychology*, 9, 35-39.
- Campbell, T. and Pettigrew, J. (2004). Testable corollaries, a conceptual error, and neural correlates of Grush's synthesis. *Behavioral and Brain Sciences*, 27, 398-400.
- Carruthers, P. (2006). Why Pretend? In S. Nichols (Ed.), *The Architecture of Imagination*. Oxford: Oxford University Press.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chalmers, D. (1999). Materialism and the Metaphysics of Modality. *Philosophy and Phenomenological Research*, 59(2), 473-496.
- Chalmers, D., & Jackson, F. (2001). Conceptual Analysis and Reductive Explanation. *The Philosophical Review*, 110(3), 315-360.
- Chapman, L.J. and Chapman, J.P. (1988). The genesis of delusions. In T. Oltermanns and B. Maher (eds.), *Delusional Beliefs*. New York: Wiley.
- Chatterjee, M. D., & Southwood, M. H. (1995). Cortical blindness and visual imagery. *Neurology*, 45, 2189-2195.
- Clayton, N. S., Bussey, T. J., & Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nat Rev Neurosci*, 4, 685-691.
- Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, 395, 272-274.
- Craver-Lemley, C., & Reeves, A. (1992). How visual imagery interferes with vision. *Psychological Review*, 99(4), 633-649.

- Currie, G. (1995). Imagination and simulation: Aesthetics meets cognitive science. In M. Davies & T. Stone (Eds.), *Mental Simulation: Evaluations and Applications*. Oxford: Basil Blackwell.
- Currie, G. and Ravenscroft, I. (1997). Mental simulation and motor imagery. *Philosophy of Science*, 64, 161-180.
- Currie, G., & Ravenscroft, I. (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Clarendon Press.
- Danckert, J., Rossetti, Y., d'Amato, T., Dalery, J., and Saoud, M. (2002). Exploring imagined movements in patients with schizophrenia. *Neuroreport*, 13, 605-609.
- Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, and Jeannerod, M. (1997). Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, 65, 71-86.
- Davidson, R. and Schwartz, G. (1977). Brain mechanisms subserving self-generated imagery: electrophysiological specificity and patterning. *Psychophysiology*, 14, 598-602.
- Davies, M., & Stone, T. (2001). Mental Simulation, Tacit Theory, and the Threat of Collapse. *Philosophical Topics*, 29, 127-173.
- Dell, G.S. (1978). Slips of the mind. In M. Paradis (ed.), *The first LACUS Forum*. Columbia, SC: Hornebeam Press.
- Dennett, D. (1981). The Nature of Images and the Introspective Trap. In N. J. Block (Ed.), *Imagery*. Cambridge, MA: MIT Press.
- Dierks, T., Linden, D.E.J., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., and Singer, W. (1999). Activation of Heschl's gyrus during auditory

- hallucinations. *Neuron*, 22, 615-622.
- Evans, C.L., McGuire, P.K., and David, A.S. (2000). Is auditory imagery defective in patients with auditory hallucinations? *Psychological Medicine*, 30, 137-148.
- Farah, M. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition*, 18, 245-272.
- Finke, R. A. (1989). *Principles of Mental Imagery*. Cambridge, MA: MIT Press.
- Finke, R. A., & Shepard, R. N. (1986). Visual functions of mental imagery. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 2, pp. 1-55). New York: Wiley-Interscience.
- Fodor, J. A. (1975). *The language of thought*. New York: Crowell.
- Fodor, J. A. (1983). *The Modularity of Mind: An Essay of Faculty Psychology*. Cambridge, MA: MIT Press.
- Fodor, J. A. (2003). *Hume Variations*. Oxford: Oxford University Press.
- Ford, J.M. and Mathalon, D.H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research*, 38, 37-46.
- Ford, J.M. and Mathalon, D.H. (2005). Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *International Journal of Psychophysiology*, 58, 179-189.
- Frith, C. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hove, UK: Lawrence Erlbaum.
- Frith, C.D. (2005). The neural basis of hallucinations and delusions. *Comptes Rendus Biologies*, 328, 169-175.

- Frith, C.D. and Done, D.J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine*, 19, 359-63.
- Gallagher, S. (2000). Self-reference and schizophrenia: a cognitive model of immunity to error through misidentification,' In D. Zahavi (ed.), *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-experience*. Amsterdam and Philadelphia: John Benjamins.
- Gallagher, S. (2004). Neurocognitive models of schizophrenia: a neurophenomenological critique. *Psychopathology*, 37, 8-19.
- Gendler, T. (2000). The puzzle of imaginative resistance. *Journal of Philosophy*, 97(2), 55.
- Gendler, T. (Ed.) (2005) *Encyclopedia of Philosophy* (2 ed., Vols. 4). Detroit: Macmillan Reference USA.
- Gendler, T. S. (2003). On the Relation between Pretense and Belief. In M. Kieran (Ed.), *Imagination, Philosophy and the Arts* (pp. 125-141). London: Routledge.
- Gendler, T. S. (2006). Imaginative Contagion. *Metaphilosophy*, 37(2), 183-203.
- Goldenberg, G., Podreka, I., Uhl, F., Steiner, M., Willmes, K., and Deecke L. (1989). Cerebral correlates of imagining colours, faces, and a map. *Neuropsychologia*, 27, 641-664.
- Goldman, A. (2006a). Imagination and Simulation in Audience Responses to Fiction. In S. Nichols (Ed.), *The Architecture of Imagination* (pp. 41-56). Oxford: Oxford University Press.
- Goldman, A. (2006b). *Simulating Minds*. Oxford: Oxford University Press.
- Goodman, N. (1976). *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis,

- IN: Hackett.
- Gordon, R. M., & Barker, J. (1994). Autism and the 'theory of mind' debate. In G. Graham & L. Stephens (Eds.), *Philosophical Psychopathology: A Book of Readings* (pp. 163-184). Cambridge, MA: MIT Press.
- Gould, L.N. (1949). Auditory hallucinations and subvocal speech: objective study in the case of schizophrenia. *Journal of Nervous and Mental Disease*, 109, 418-427.
- Graham, G. and Stephens, G.L. (2000). *When Self-Consciousness Breaks*. Cambridge, MA: MIT Press.
- Green, P. and Preston, M. (1981). Reinforcement of vocal correlates of auditory hallucinations by auditory feedback: case study. *British Journal of Psychiatry*, 139, 204-208.
- Greenberg, D. L., Eacott, M. J., Brechin, D., & Rubin, D. C. (2005). Visual memory loss and autobiographical amnesia: a case study. *Neuropsychologia*, 43(10), 1493-1502.
- Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27, 377-442.
- Hadwin, J., & Bruins, J. (1998). *The role of counterfactual thought in children's judgements of sadness*. Paper presented at the XVth Biennial Meeting of the International Society for the Study of Behavioral Development.
- Haggard, P. (2003). Conscious awareness of intention and of action. In J. Roessler, J. and N. Eilan, (eds.), *Agency and Self-awareness*. Oxford: Oxford University Press.

- Harris, P. L. (2000). *The Work of the Imagination*. Oxford: Blackwell.
- Harris, P. L. (2001). The Veridicality Assumption. *Mind and Language*, 16(3), 247-262.
- Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proc Natl Acad Sci U S A*, 104, 1726-1731.
- Heal, J. (1994). Simulation vs. Theory Theory: What Is at Issue? In C. Peacocke (Ed.), *Objectivity, Simulation and the Unity of Consciousness* (pp. 129-144). Oxford: Oxford University Press.
- Heal, J. (1998). Co-cognition and off-line simulation: two ways of understanding the simulation approach. *Mind and Language*, 14, 477-498.
- Heck, R. G. (2000). Nonconceptual content and the space of reasons. *The Philosophical Review*, 109, 483-523.
- Helmholtz, H. (1866). *Handbuch der Physiologischen Optik*. Leipzig: Voss.
- Hill, C. (1997). Imaginability, conceivability, possibility and the mind-body problem. *Philosophical Studies*, 87, 61-85.
- Hill, C., & McLaughlin, B. P. (1999). There are fewer things in reality than are dreamt of in Chalmers's Philosophy. *Philosophy and Phenomenological Research*, 59(2), 445-454.
- Jackson, F. (1987). *Conditionals*. New York, NY: Basil Blackwell.
- Janssen, J. J., & Sheikh, A. A. (1994). Enhancing Athletic Performance Through Imagery: An Overview. In A. Sheikh & E. Korn (Eds.), *Imagery in Sports and Physical Performance* (pp. 1-22). New York: Baywood Publishing Company.
- Jeannerod, M., & Pacherie, E. (2004). Agency, simulation and self-identification. *Mind and Language*, 19, 113-146.

- Jones, S.R. and Fernyhough, C. (2006). Thought as action: inner speech, self monitoring, and auditory verbal hallucinations. *Consciousness and Cognition* (forthcoming).
- Kapur, N. (1993). Focal retrograde amnesia in neurological disease: a critical review. *Cortex*, 29, 217-234.
- Kosslyn, S.M. (1978). Measuring the visual angle of the mind's eye. *Cognitive Psychology*, 10, 356-389.
- Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press.
- Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Thompson, W. L., et al. (1999). The role of area 17 in visual imagery: Convergent evidence from PET and rTMS. *Science*, 2(5411), 167-170.
- Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, 129, 723-746.
- Kosslyn, S. M., Thompson, W. L., & Alpert, N. M. (1997). Neural systems shared by visual imagery and visual perception: A positron emission tomography study. *Neuro-Image*, 6, 320-334.
- Kripke, S. A. (1980). *Naming and necessity*. Cambridge, Mass.: Harvard University Press.
- Kripke, S. A. (1982). *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.
- Lang, P. (1984). Cognition and Emotion: Concept and Action. In C. Izard, J. Kagan & R. Zajoncs (Eds.), *Emotions, Cogntion and Behavior* (pp. 192-226). Cambridge: Cambridge University Press.

- Leevers, H. J., & Harris, P. L. (2000). Counterfactual Syllogistic Reasoning in Normal 4-Year-Olds, Children with Learning Disabilities, and Children with Autism. *Journal of Experimental Child Psychology*, 76, 64-87.
- Leslie, A. (1987). Pretense and representation: the origins of "theory of mind". *Psychological Review*, 94, 412-426.
- Leslie, A. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, 50, 211-238.
- Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.
- Lewis, D. (1980). Veridical Hallucination and Prosthetic Vision. *Australasian Journal of Philosophy*, 58, 239-249.
- Lindner, A., Their, P., Kirchr, T.T.J., Haarmeier, T., Leube, D.T. (2005). Disorders of agency in schizophrenia correlate with an inability to compensate for the sensory consequences of actions. *Current Biology*, 15, 1119-1124.
- Lormand, E. (2007). Phenomenal Impressions. In T. Gendler & J. Hawthorne (Eds.), *Perceptual Experience* (pp. 317-353). Oxford Oxford University Press.
- Maher, B.A. (1974). Delusional thinking and perceptual disorder. *Journal of Individual Psychology*, 30, 98-113.
- Margo, A., Hemsley, D.R., and Slade, P.D. (1981). The effects of varying auditory input on schizophrenic hallucinations. *British Journal of Psychiatry*, 139, 122-127.
- Martin, M. G. F. (2002). The Transparency of Experience. *Mind and Language*, 17(4), 376-425.

- McDowell, J. (1994). *Mind and World*. Cambridge, MA: Harvard University Press.
- McGinn, C. (2000). *The Mysterious Flame: Conscious Minds in a Material World*. New York, NY: Basic books.
- McGinn, C. (2004). *Mindsight : image, dream, meaning*. Cambridge, Mass.: Harvard University Press.
- McLaughlin, B. P., & Tye, M. (1998). Is Content-Externalism Compatible with Privileged Access? *The Philosophical Review*, 107(3), 349-380.
- Meehan, M. Insko, B., Whitton, M. and Brooks, F.P. (2002). Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics*, 21, 645-652.
- Mel, B.W. (1988). MURPHY: A robot that learns by doing. In D.Z. Anderson (ed.), *Neural information processing systems*, American Institute of Physics.
- Mellor, C.H. (1970). First rank symptoms of schizophrenia. *British Journal of Psychology*, 117, 15-23.
- Miall, R.C., Weir, D.J., Wolpert, D.M., and Stein, R.C. (1993). Is the cerebellum a smith predictor? *Journal of Motor Behavior*, 25, 203-216.
- Nayani, T.H., David, A.S. (1996). The auditory hallucination: a phenomenological survey. *Psychological Medicine*, 26, 177-189.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435-450.
- Nelson, H.E., Thrasher, S., and Barnes, T.R.E. (1991). Practical ways of alleviating auditory hallucinations. *British Medical Journal*, 302, 327.
- Nichols, S. (2004). Imagining and Believing: The Promise of a Single Code. *Journal of Aesthetics and Art Criticism*, 62, 129-139.

- Nichols, S. (2006a). Introduction. In S. Nichols (Ed.), *The Architecture of the Imagination* (pp. 1-18). Oxford: Oxford University Press.
- Nichols, S. (2006b). Just the Imagination: Why Imagining Doesn't Behave Like Believing. *Mind and Language*, 21, 459-474.
- Nichols, S., & Stich, S. (2000). A cognitive theory of pretense. *Cognition*, 74, 115-147.
- Nichols, S., & Stich, S. (2003). *Mindreading*. Oxford: Oxford University Press.
- Noordhof, P. (2002). Imagining objects and imagining experiences. *Mind and Language*, 17(4), 426-455.
- Papineau, D. (2002). *Thinking About Consciousness*. Oxford: Oxford University Press.
- Papineau, D. (2007). Kripke's Proof is Ad Hominem Not Two-Dimensional. *Philosophical Perspectives*, 21(1), 475-494.
- Peacocke, C. (1985). Imagination, Possibility and Experience. In J. Foster & H. Robinson (Eds.), *Essays on Berkeley: a Teracentennial Celebration*. Oxford: Clarendon Press.
- Perky, C. W. (1910). An experimental study of imagination. *The American Journal of Psychology*, 21(3), 422-452.
- Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.
- Peterson, D. M., & Bowler, D. M. (1996). *Subtractive reasoning and false belief understanding in autistic, SLD and non-handicapped children*. Paper presented at the British Psychological Society, Developmental Psychology Section Annual Conference.
- Peterson, D. M., & Bowler, D. M. (2000). Counterfactual reasoning and false belief understanding in children with autism. *autism*, 4(4), 391-405.
- Prinz, J. (2007). Beyond Appearances: The Content of Sensation and Perception *Perceptual*

- Experience* (pp. 434-460). Oxford: Oxford University Press.
- Prothero, J.D. (1993). The treatment of akinesia using virtual images. *Master's thesis*, Industrial Engineering, University of Washington.
- Proust, J. (2003). Thinking of oneself as the same *Consciousness and Cognition*, 12, 495-509.
- Proust, J. (2006). Agency in schizophrenia from a control theory viewpoint. In N. Sebanz & W. Prinz (Eds.), *Disorders of Volition*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25(2), 157-237.
- Roland, P. and Friberg, L. (1985). Localization of cortical areas activated by thinking. *Journal of Neurophysiology*, 53, 1219-1243.
- Rothbaum, B., Hodges, L, Alarcon, R., Ready, D., Shahar, F., Graap, K., Pair, J., Herbert, P., Gotz, D., Wills, B., and Baltzell, D. (1999). Virtual reality exposure therapy for PTSD Vietnam veterans: a case study. *Journal of Traumatic Stress*, 12, 263–71.
- Rubin, D. C., & Greenberg, D. L. (1998). Visual memory-deficit amnesia: A distinct amnesiac presentation and etiology. *Proc Natl Acad Sci U S A*, 95, 5413-5416.
- Rubin, D. C., & Greenberg, D. L. (2003). The Neuropsychology of Autobiographical Memory. *Cortex*, 39, 687-728.
- Sanchez-Vivesm M. and Slater, M. (2005). From presence towards consciousness. *Nature Reviews Neuroscience*, 6, 332.
- Sass, L.A. and Parnas, J. (2003). Schizophrenia, consciousness, and the self. *Schizophrenia Bulletin*, 29, 427-444.

- Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nat Rev Neurosci*, 8(9), 657-661.
- Schneider, K. (1959). *Clinical Psychopathology*. New York: Grune and Stratton.
- Schroeder, T., & Matheson, C. (2006). Imagination and Emotion. In S. Nichols (Ed.), *The Architecture of the Imagination* (pp. 19-40). Oxford: Oxford University Press.
- Scott, F. J. (1999). 'If pigs could fly': A test of counterfactual reasoning and pretense in children with autism. *British Journal of Developmental Psychology*, 17, 349-362.
- Searle, J. (1983). *Intentionality: an essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Searle, J. (1991). Response: Reference and Intentionality. In E. Lepore & R. Van Gulick (Eds.), *John Searle and his Critics* (pp. 227-237). Oxford: Blackwell.
- Segal, S. J. (1971). Processing the Stimulus in Imagery and Perception. In S. J. Segal (Ed.), *Imagery: Current Cognitive Approaches* (pp. 73-100). New York: Academic Press.
- Shergill, S.S., Samson, G., Bays, P.M., Frith, C.D., Wolpert, D.M. (2005). Evidence for sensory prediction deficits in schizophrenia. *American Journal of Psychiatry*, 162, 2384-2386.
- Siegel, S. (2007). Which Properties are Represented in Perception? In T. Gendler & J. Hawthorne (Eds.), *Perceptual Experience* (pp. 481-503). Oxford: Oxford University Press.
- Sirigu, A., Cohen, L., Duhamel, J. R., Pillon, B., Dubois, B., Agid, Y., et al. (1995). Congruent unilateral impairments for real and imagined hand movements. *NeuroReport*, 6, 997-1001.
- Slotnick, S., Thompson, W., & Kosslyn, S. M. (2005). Visual mental imagery induces

- retinotopically organized activation of early visual areas. *Cerebral Cortex*, 15, 1570-1583.
- Spence, S. (1996). Free will in the light of neuroplasticity. *Philosophy, Psychiatry, and Psychology*, 3, 75-90.
- Sperry, R.W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *The Journal of Comparative and Physiological Psychology*, 43, 482-489.
- Sorenson, R. (1996). Modal Bloopers: Why Believable Impossibilities are Necessary. *American Philosophical Quarterly*, 33(1), 247-261.
- Stephane, M., Barton, S. and Boutros, S. (2001). Auditory verbal hallucinations and dysfunction of the neural substrates of speech. *Schizophrenia Research*, 50, 61-78.
- Stich, S., & Nichols, S. (1997). Cognitive Penetrability, Rationality and Restricted Simulation. *Mind and Language*, 12(3/4), 297-326.
- Stich, S., & Nichols, S. (1998). Rethinking Co-Cognition: A Reply to Heal. *Mind and Language*, 13(4), 499-512.
- Stoljar, D. (2006). *Ignorance and Imagination*. Oxford: Oxford University Press.
- Strawson, G. (2003). Mental ballistics or the involuntariness of spontaneity. *Proceedings of the Aristotelian Society*, 103, 227-56.
- Suinn, R. M. (1994). Visualization in Sports. In A. Sheikh & E. Korn (Eds.), *Imagery in Sports and Physical Performance* (pp. 23-42). New York: Baywood Publishing Company.
- Thomas, N. (2007). Mental Imagery. from Edward Zalta:

- <http://plato.stanford.edu/archives/win2007/entries/mental-imagery/>
- Tulving, E. (2005). Episodic memory and Autonoesis: Uniquely Human? In H. S. Terrace & J. Metcalfe (Eds.), *The Missing Link in Cognition* (pp. 3-56). Oxford: Oxford University Press.
- Tye, M. (1988). The Picture Theory of Mental Images. *The Philosophical Review*, 97(4), 497-520.
- Tye, M. (1991). *The imagery debate*. Cambridge, Mass.: MIT Press.
- Usuh, M., Catena, E., Arman, S. and Slater, M. (2000). Using presence questionnaires in reality. *Presence-Teleoperators and Virtual Environments*, 9, 497-503.
- Velleman, J. D. (2000). *The Possibility of Practical Reasoning*. Oxford: Oxford University Press.
- Vosgerau, G. and Newen, A. (2007). Thoughts, motor actions, and the self. *Mind and Language*, 22, 22-43.
- Walton, K. L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, MA: Harvard University Press.
- Weinberg, J., & Meskin, A. (2006a). Imagine That! In M. Kieran (Ed.), *Contemporary Debates in Aesthetics and the Philosophy of Art* (pp. 222-235). Oxford: Blackwell.
- Weinberg, J., & Meskin, A. (2006b). Puzzling Over the Imagination: Philosophical Problems, Architectural Solutions. In S. Nichols (Ed.), *The Architecture of Imagination* (pp. 175-204). Oxford: Oxford University Press.
- Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated

- abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, 9(1), 11-29.
- Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: B. Blackwell.
- Wittgenstein, L. (1969). *The Blue and Brown Books*. New York: Harper.
- Wright, C. (2002). The Conceivability of Naturalism. In T. Gendler & J. Hawthorne (Eds.), *Conceivability and Possibility* (pp. 401-440). Oxford: Oxford University Press.
- von Holst and Mittelstadt, H. (1950/1973). Das Reafferenzprinzip. *Naturwissenschaften*, 37, 464-76.
- English translation, 1973: The reafference principle. In R. Martin (trans), *The Behavioral Physiology of Animals and Man. The Collected Papers of Erich von Holst*, trans. R. Martin. Miami, FL: University of Miami Press.
- Vygotsky, L.S. (1934/1987). *Thought and language*. Cambridge MA: MIT Press.
- Original work published in 1934.
- Wann, J. (1996). Virtual reality environments for rehabilitation of perceptual-motor disorders following stroke. *Proceedings of the First European Conference on Disability, Virtual Reality and Associated Technologies*, 1, 233-238.
- Weiskrantz, L., Elliot, J., and Darlington, C. (1971). Preliminary observations of tickling oneself. *Nature*, 230, 598-99.
- Wittgenstein, L. (1953). *Philosophical Investigations*. 2nd ed. trans. G.E.M. Anscombe. Oxford: Blackwell.
- Wolpert, D. and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience Supplement*, 3, 1212-1217.

- Wolpert, D., Miall, R.C., and Kawato, M. (1998). Internal models in the cerebellum. *TRENDS in Cognitive Science*, 2, 338-347.
- Yablo, S. (1999). Concepts and Consciousness. *Philosophy and Phenomenological Research*, 59(2), 455-463.
- Yablo, S. (2006). No fool's cold: notes on illusions of possibility. In J. Macià & M. García-Carpintero (Eds.), *Two-Dimensional Semantics* (pp. 327-380). Oxford: Oxford University Press.
- Zaretsky, M. and Rowell, C.H.F. (1979). Saccadic suppression by corollary discharge in the locust. *Nature*, 280, 583-85.
- Zimmons, P. and Panter, A. (2003). The influence of rendering quality on Presence and task performance in a virtual environment. In *Proceedings of the IEEE Virtual Reality Conference 2003*. IEEE Computer Society, 293.