A Pluralistic Account of Propositional Imagination

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Michael Joseph Ferreira

Graduate Program in Philosophy

The Ohio State University

2014

Dissertation Committee:

Richard Samuels, Advisor

Timothy Schroeder

Declan Smithies

Abstract

We exercise the propositional imagination whenever we imagine that *p* – e.g. that *it's snowing outside*, that *Othello murders Desdemona*, or that *cats are actually Martian-controlled robots*. Here I aim to sketch a pluralistic account of propositional imagination, according to which the cognitive phenomena associated with imagination are underpinned by multiple kinds of psychological state. I begin by presenting the default cognitive account of propositional imagination. What I call *the new cognitive theory* has played a central role in displacing early attempts in developmental psychology to link pretend play in toddlers to an early capacity to reason about the unobservable psychological states of oneself and others. Roughly put, the new cognitive theory casts imagination as a *distinct cognitive attitude*, yet one that is compositionally akin to belief. I argue, however, that there's a deep explanatory tension in this account's core commitments. In particular, the view faces *the asymmetry challenge*; for, the vehicles of imagination are cast as so very similar to those of belief that there seems to be little reason to suppose that they should play the robustly distinct functional role that the theory demands. Next, I evaluate an emerging alternative approach – *the single attitude account* – which assimilates the mechanisms and vehicles of propositional imagination to those of counterfactual reasoning generally. I argue that the alternative approach fails to accommodate important tracts of data surrounding our consumption and production of

fictions. In the penultimate chapter, I consider how these two accounts of imagination propose to understand the architecture of pretense. One important, unresolved issue here surrounds the question of whether children require recourse to *metacognition* – i.e. beliefs about imagination, and perhaps other mental states – in order to recognize and engage in pretense. I argue that – in spite of the suggestions by the proponents of both new cognitivism and the single attitude approach – no satisfactory alternative to the metacognitive approach is forthcoming. Hence, we should assume that pretense (at least pretense *recognition*) requires metacognitive states. Finally, I sketch a pluralistic account of imagination, arguing that the psychological vehicles of imagination are diverse, populating at least three distinct psychological categories: counterfactual elaboration, bare imaginings (i.e. mere entertaining that *p*), and fictionalized attitudes akin to the metarepresentational pretense states initially put forward in metacognitive accounts of pretense. With little cost, a robust pluralism about the mechanisms and vehicles of imagination at once promises to capture much of the data unified theories fail to capture, while also avoiding the asymmetry challenge I raise for the new cognitive theory.

For Reeni & Joe

Acknowledgments

I would like to thank Richard Samuels for many helpful hours of discussion regarding the issues considered in this dissertation, and for comments on countless drafts. I'd like also to thank Declan Smithies, Timothy Schroeder, Shaun Nichols, Peter Langland-Hassan, Amy Kind, Neil van Leeuwen, Daniel Pearlberg, Justin D'Arms, Gualtiero Piccinini, Salvatore Florio, and Daniel Wilkenfeld for their valuable comments and criticisms.

Vita

June 1995 .....Saint Raphael Academy, Pawtucket, RI

1999.....B.A. University of San Francisco

2006.....M.A. University of Missouri – St. Louis

2006 to present.....Graduate Teaching Associate, Department of Philosophy, The Ohio State University

Publications

"Why don't concepts constitute a natural kind?" (with Richard Samuels), *Behavioral and Brain Sciences* 33 (2010): 222-223.

Fields of Study

Major Field: Philosophy

Table of Contents

Preliminaries

'But there is a different…use of the term 'imagination' under which one needs to distinguish between imagery and imagination. In this sense…one can imagine a state of affairs without having any imagery of it. It is this usage that is the one which we will now target.' (B. Gaut, 'Creativity and Imagination,' 2003)

This dissertation is about the psychological vehicles of imagination. I examine some *prima facie* plausible accounts of the psychological vehicles subserving our capacity to represent propositional contents under the peculiar light of imagination. Following some primary players in the philosophy and cognitive science of imagination, let's call the target phenomenon *the propositional imagination*.

Prima facie, we engage in propositional imagining when we imagine that *such-and-such is the case*. Thus, the propositional imagination seems to be at work whenever we imagine that *p*, where *p* stands for any representable proposition – e.g. *that cats are alien robots* or that *I win the* lottery. I think that in spite of some very productive recent scholarly efforts to understand what's going on (psychologically) in such cases, we still don't have a very good grasp of the phenomenon of imagining that *p*. It is the broad aim of this dissertation is to tighten that grasp. I mean to do so by arguing that recent efforts to provide a unified cognitive theory of propositional imagination have failed. They have

failed, in sum, because there is no unified phenomenon of which to give an account. In short, what scholars have been calling the 'propositional imagination' – assuming all the while that the term is univocal in its reference to the members of a special psychological kind – is really a diverse array of underlying representational phenomena.

It is worth observing that there's much at stake, since there is a venerable tradition of romanticizing the imagination, on the one hand, and making heavy theoretical use of it, on the other. On the whole, intellectuals of diverse stripes (and the psychologizing folk, for that matter) have tended to laude the imagination as a kind of rogue, independent, and sometimes superior mental faculty.[1]

The imagination, in contrast with whatever faculties underwrite the production and maintenance of our work-a-day beliefs, has been cast as a free, unbounded cognitive capacity, subject (in principle) to few limitations. Imagination is thus frequently associated with creativity, originality, ingenuity, artistic achievement, and scientific discovery. Given its associations and credits, the imagination is clearly expected to play an important role in the explanation of many of our species' most impressive and distinctive attributes.

Enthusiasm over the nature and role of imagination is not new. As early as Aristotle (Hicks, 1907) the imagination was awarded a central role in explaining the workings of the mind. Aristotle believed the imagination necessary for the very possibility of cognition; for – *ex hypothesi* – imaginings furnished the mind with the

[1] See, for instance, S. Johnson's (1751) *Rambler* (no. 125) for a characteristically romantic view of imagination: 'Imagination, a licentious and vagrant faculty, unsusceptible of limitations and impatient of restraint, has always endeavoured to baffle the logician, to perplex the confines of distinction, and burst the enclosures of regularity.'

images crucial for constructing the psychological vehicles of mental representation. For Kant (1934), the imagination played a critical role in synthesizing the deliverances of perception and those of the intellect, thus playing a primary role in structuring our knowledge and our very experience of reality.

In Hume's works (e.g. Hume, 1999), we find the imagination underpinning a wide variety of cognitive functions, e.g. the combination, analysis, and reproduction of ideas. For Hume, as for Aristotle and Kant, the imagination was thus a critical source of knowledge. Hume, moreover, famously took imagination to be the arbiter of logical possibility.

Descartes (1998) stands out as an important figure who (at least sometimes) *downplayed* the cognitive role of imagination. But there's a potentially telling explanation for this. Descartes tended to reserve the term 'imagination' for reference to mental imagery (and the faculty responsible for conjuring mental imagery). And – unlike Aristotle, Kant, and Hume – Descartes did not think that mental imagery played any essential role in our cognitive lives. Descartes thus drew a sharp distinction between imagining (*qua* conjuring imagery) and the having of cognitive states which many others (especially recently) have been wont to associate with imagination, and with propositional imagination, in particular. For instance, states we ordinarily pick out with terms like 'entertaining', 'conceiving', 'supposing', 'pretending' and 'hypothesizing' are,

today, usually assumed to fall under, or to be some close relation to, the propositional imagination.[2]

Since efforts at giving a unified (cognitive) account of propositional imagining have tended to assume that the latter terms are at least close relatives of imagination, it's arguable that *prima facie* disputes between Descartes and other historical figures regarding the centrality of imagination in cognition are largely terminological. Any *prima facie* dispute would seem only to be over whether we *label* the diverse range of propositional attitudes we pick out in ordinary language with term 'imagination.'

Historical terminological issues to one side, scholarly enthusiasm over the propositional imagination is presently high and on the rise. Propositional imagining has been invoked in explanations of phenomena as diverse as, *inter alia*, creativity (Gaut 2003; Boden, 2004; Carruthers 2006), pretend play in children and adults (Leslie 1987; Nichols and Stich 2003), the capacity to predict and explain apparently purposeful behavior (Leslie 1987, 1994; Gordon 1986; Gordon and Barker 1994; Currie and Ravenscroft 2002; Nichols and Stich 2003; Goldman 2006), planning (Harris 1993; Goldman 1992), counterfactual reasoning (Goldman 1992; Harris 2000), modal reasoning (Lewis 1986; Rosen 1990; Sidelle 2002; Gendler and Hawthorne 2002), dreaming and visual imagery (Currie 1995; McGinn 2004; Ichikawa, 2009), strategy testing (Currie 1995b), and our cognitive and emotional engagement with fictional works (Currie 1995a; Walton 1990, 1997; Goldman 2006; Nichols 2004a, 2004b, 2006a, 2006b).

[2] In fact, Descartes apparently may have had a more nuanced view of imagination – perhaps to include something like a notion of propositional, or intellectual, imagination (see Sepper, 1996).

But despite the imagination's widely and historically acknowledged importance, and despite recent trends in philosophical and psychological research, there are remarkably few facts about the propositional imagination, *per se*, to grasp and theorize around. Hence, the imagination remains today – to borrow a vivid phrase from Shaun Nichols – 'one of the darker faculties of the human mind'(Nichols 2006b).

Needless to say, given the array of theoretical contexts in which the imagination is invoked, it's clear that a full accounting of a very long list of important phenomena ultimately hinges on an adequate theory of propositional imagination.

1. *What is the Propositional Imagination?*

The first thing to notice is that what I'm taking for the target phenomenon here is picked out by a term of art – i.e. 'propositional imagination' – which has come to occupy a proprietary place in the philosophy of cognitive science. It's important, moreover, to bear in mind that the term is a designation which has evolved over the course of the last thirty years.

To begin to get purchase on the putative phenomenon, it'll be useful to draw out the intuitive contrast attributed to Descartes above: This is the contrast between the conjuring of mental imagery, on one hand, and the kinds of mental tokenings which underpin propositional attitudes we might pick out with terms like 'entertaining', 'conceiving', or 'supposing', on the other.

Conjuring imagery, or what we might call 'perceptual' imagining, allows us to enter into qualitatively rich simulations of perceptual experiences. For example, we can (perceptually) imagine *the way the Mona Lisa looks*, *the way an oboe sounds in some particular concert hall*, or *the way cranberry sauce tastes*. We can imagine such things even when we are far removed from the experience of the relevant phenomena. In such cases, moreover, the perceptual imagery represented is *individuating* for the token imagining, and we might say that the images are what constitute the content of the imagining.

By contrast, we might also entertain a *propositional* content. We might, for instance, entertain propositions like that *it's snowing on Mount Everest*, that *Othello murders Desdemona*, or that *a runaway train is about to plow through a crowd of civilians*. When we do so, we represent a proposition *p,* but without (necessarily) believing or (necessarily) desiring the state of affairs described by *p*.

Moreover, unlike perceptual imaginings, any imagery associated with the entertained content, *p*, is inessential to the individuation of the token representation. *Prima facie*, for instance, you and I needn't have any particular imagery conjured in order for it to be true that we both are entertaining that *Othello is a jealous man*. And also, our mental images of Othello may vary greatly, in spite of the fact that we both entertain this content.

2. *Setting the Agenda: Classical Metarepresentationalism About Pretense*

In his seminal paper on the psychology of pretense, Alan Leslie vividly outlined the sorts of questions which continue to drive the scholarly debates on pretense and imagination (Leslie, 1987, p. 412):

> 'How is it possible for a child to think about a banana as if it were a telephone, a lump of plastic as if it were alive, or an empty dish as if it contained soap? If a representational system is developing, how can its semantic relations tolerate distortion in these more or less arbitrary ways? Indeed, how is it possible that young children can disregard or distort reality in any way and to any degree at all? Why does pretending not undermine their representational system and bring it crashing down?'

Since Leslie effectively set the agenda for research in the theory of propositional imagining, I'll provide a broad overview of his view here in my introduction.

For his part, Leslie addresses the question by imputing a mechanism for metacognition, or metarepresentation, to developing pretenders. The core idea behind Leslie's variety of metarepresentationalism is to exploit representational complexity in two ways. It's proposed that the imaginational representations subserving pretense are

beliefs which bear (*i*) proprietary structural properties, and (perhaps more contentiously) (*ii*) proprietary contents.

On this account, the pretense representations subserving pretense detection are identified with internal representations of the form **Agent – Informational Relation – *'p'***. The informational relation, for its part, is captured by a psychologistic concept PRETEND, and the quotes around ***'p'*** mark additional representational structure.[3] For their part, the additional structural features serve to indicate (within the system) that ***p*** is not the object of belief for the purposes of downstream processing. Hence, *p* is effectively *decoupled* from the input/output relations characteristic of the belief that *p*. For instance, my pretending that the recess bell is ringing involves tokening the representation I – PRETEND – 'THAT THE RECESS BELL IS RINGING'. Pretense recognition proceeds by tokening similar representations, differing in that reference to a 3rd party occurs in the **Agent** position.[4]

Architecturally speaking, Leslie posits three proprietary processes realized in a cognitive mechanism called the *Decoupler*. The *Decoupler* then decomposes into three submechanisms, the collective workings of which are supposed to account for the production of imaginational representations, their differential treatment, while allowing for their similar treatment wherever necessary. The *Decoupler* houses, the *Expression*

[3] Notice, also, that the concept PRETEND is supposed to be primitive, atomistic, and innate. Thus, the concept is not (necessarily) semantically similar to whatever the folk psychological term 'pretend' expresses.

[4] A major feature of Leslie's account is his proposed isomorphism between properties of imaginational representations and properties of sentences which contain mental state terms. The putative isomorphism is pressed into service as evidence that pretense behavior belies the onset of a capacity to understand cognition, generally. In effect, the proposal is that 'pretense is an early manifestation of what has been called *theory of mind*'(Leslie 1987, p. 416; Premack & Woodruff 1978).

Raiser, the *Manipulator*, and the *Interpreter*, and *both* pretense *recognition* and *engagement* proceeds *via* the activation of imaginational representations as follows.

Let me give a rough illustration of how this is supposed to work. Take a case where it is pretended that *a banana is a telephone*. The *Expression Raiser* decouples the representation THAT [THIS] IS A BANANA, outputting the altered (quoted) representation 'THAT [THIS] IS A BANANA'. This representation is, by hypothesis, quarantined from the semantic relations of the former, and is primed for manipulation as a purely formal object. The *Manipulator* then transforms the decoupled expression to 'THAT [THIS] BANANA, IT IS A TELEPHONE'. Finally, the *Interpreter* delineates the relations between the decoupled representation 'BANANA' and primary representation BANANA. Presumably this involves tracking the salient and conspicuous ways in which the banana is naturally analogous – and *made to be* analogous – to a telephone.

For present purposes, we needn't be concerned with any further details here. However it is noteworthy that while 'PRETEND' here refers to a proprietary informational relation between an agent and a decoupled content, Leslie also notes that fully explaining pretense *recognition* requires more than the mechanism described. The process so far described could apply *only* to solitary pretense. But the final result of the *Decoupler's* operations is supposed to be the production of a complete imaginational representation of the form THAT AGENT – PRETEND(S) – 'THAT [THIS] BANANA, IT IS A TELEPHONE'. Hence, in addition to the capacity for decoupling, the young imaginer who has achieved a

competence in pretense recognition also must have a primitive capacity to attribute mental states to others.

The classical metarepresentationalist account has some *prima facie* plausibility. It describes a mechanism capable of both engaging and recognizing pretense behavior. It also gives a rough account of the early onset of imaginational thinking, linking imagination to the capacity to understand minds in general. Furthermore, it provides for an account of why belief, on the one hand, and why pretense and imagination, on the other, should be functionally similar: For, they issue in the psychological vehicles of the same fundamental variety. But it also has a principled means of explaining asymmetries across belief and imagination/pretense: The former are functionally dissimilar from the latter because of their distinctive compositional features.

3. *A Prevalent Conception of Propositional Imagination*

Metacognitivism about pretense and imagination has fallen mostly out favor these days. What we have seen, at least in the philosophical literature, is the evolution of common widespread conception of imagination, which is lately come to be manifest in a powerful new cognitive theory of propositional imagination (Nichols and Stich, 2003).

The thread this new theory picks up and elaborates upon revolves around the idea that token entertainings of the relevant sort are the internal, mental analogues of non-assertoric utterances (Scruton, 1974; McGinn, 2004). Notice, for example, that we can

give *verbal* expression to propositional contents more or less at will – e.g. I can utter: 'Washington D.C. is the capital of Rhode Island.' But when I say it, it doesn't mean that I'm declaring that I believe it to be true. Nevertheless, the utterance has truth conditions, and I know what it would mean for it to be true. Moreover, my utterance presumably represents the same (false) state of affairs that it would in the mouth of someone making the (false) assertion that D.C. is the capital of RI.

Intuitively, propositional imaginings (*qua* entertainings) are something very like non-assertorical utterances. There are evident, *prima facie* similarities between merely uttering *v*. asserting '*p'* on the one hand, and entertaining *vs*. believing that *p*, on the other. But, for the simple range of sentence types there are, it's evident that that analogy only carries us so far. No one, I presume, would want to say that what makes an attitude a state of entertaining (or imagining, for that matter) is *merely* the criterion of being non-assertoric.

But there are plenty of related and similarly intuitive metaphors which can, and have, been used to illuminate the conception of imagination as mere entertaining, or representation without alethic commitment. Imagination has been cast variously in terms of make-believe, games, pretense, and representations of possibilities or fictions (Walton, 1990). And token states of imagination have been characterized as decoupled beliefs, pretend beliefs, bracketed beliefs, and simulated beliefs (Leslie, 1987; Gordon, 1986; Goldman, 1992; Harris, 2000). Nevertheless, notice that the effect of each of these ways of talking about imagination parallels the effect of the utterance/assertion analogy:

Imagination is intuitively a lot like belief, minus the alethic and epistemic commitments which (necessarily) accompany the latter.

Given the *prima facie* problems with trying to categorize imaginings as (simply) a class of non-assertorical attitude, I propose to try to capture the thread running through the metaphors rehearsed under the auspices of a more general claim about imagination: Call it the *Propositional Attitude Conception (PAC)*:

> *PAC*: To imagine that *p* is to have a propositional attitude which is irreducible to any other attitude or collection of attitudes.

PAC describes a class of propositional attitude that lacks many of the essential and defining qualities of the psychological attitudes that we frequently reference in our attributions of mental states to others and ourselves.

I think that *PAC*, strictly read, tracks something true about propositional imagining. I believe this because the letter of *PAC* is consistent with my view that the propositional imagination decomposes into several representational vehicles, each of which involves the representation of a content *p*, but none of which is identifiable with the belief that *p*, full stop. However, as it's been construed in recent efforts (more of which below), *PAC* lends itself to being interpreted as (or as supporting) a less innocuous conception of propositional imagination. This less innocuous conception assumes that, since imagining (or entertaining) that *p* implies neither believing, nor desiring, nor

intending, etc.), therefore: Imagination must be subserved by an ontologically distinct psychological vehicle. Under this view, the distinctiveness of imagining is assumed to warrant its allocation to its own psychological kind, individuated in accord with a unique psychological profile. I think it's a mistake to think about the role of mental entertainings in this way. To see why, let me relate a brief, autobiographical vignette which I think illustrates the thrust of this consensus view, while orienting the reader to the arguments in the chapters that follow.

Here's my story: As a boy of around 12 I delivered *The Providence Journal* and *The Evening Bulletin* along several paper routes around the East Side of Providence. The routes included several so-called 'honor boxes,' which I was responsible for stocking each day. Back when I delivered the paper, it cost 35¢ per issue, and depositing 35¢ would thus open the box. However – once open – customers had full access to however many newspapers remained inside. Honor boxes were thus so-called, I take it, because each customer was expected to be honorable in their transaction, and to honor the rule that they retrieve only a single newspaper for every 35¢ deposited. Needless to say, not every customer did the honorable thing.

And I frequently encountered an insidious practice, a description of which will oddly enough serve as a lead into my investigation in contemporary imagination theory: Instead of paying the usual quarter and dime, some customers would deposit two round pieces of metal into the coin slot, the one the same shape, size, and weight as a quarter,

and the other the same as a dime. My father and I called these quarter and dime counterparts 'slugs'.

Now here's the point of the story: Notice that slugs played roughly the same causal role as quarters and dimes. Specifically, slugs would trigger the release of the locking mechanism, enabling the honor box door to be opened. And the honor boxes evidently had no way of responding differentially to slugs and genuine coins. For: Slugs and coins apparently shared a set of causally relevant properties with respect to honor box functioning.

Slugs help – I think – to illustrate the way a majority of philosophers of mind and cognitive science have come to understand the mechanisms and vehicles underpinning propositional imagination. If, for the moment, we can liken *genuine* quarters (and dimes) to token beliefs, then – according to the default account of imagination – *we can liken token imaginings to slugs*. Just as coins and slugs share a set of causally relevant properties, so beliefs and imaginings are supposed to share a set of causally relevant psychological properties. For instance, they're supposed to share in content, neural substrates, and syntax (if it should turn out that it's appropriate to speak of the 'syntax' of mental representation). This, in turn, is supposed to help explain, *inter alia*, the seamless inferential liaisons between belief and imagination, and why we respond emotionally to the contents of imagination. In effect, those cognitive mechanisms which receive input from belief and imagination respond similarly to both representational kinds – for, as with slugs and genuine coin, beliefs and imaginings share an important set of causally relevant properties.

Now, of course, believing that *p* is one thing, and imagining that *p* is entirely something else. No one, I take it, would contest this intuitively obvious fact. And hence the simile I align the consensus view with: *Imaginings are like slugs*. As slugs are not real money, so the contents of imaginings are not (necessarily) simple propositional models of reality. Nevertheless, according to the default view, if you were to look at a token belief with the content *p* and a token imagining with the same content (*side-by-side*, as it were), they'd look pretty much exactly the same. To stretch the slug/coin analogy to (and perhaps beyond) its limit, the belief and the imagining would have the same shape, size, and weight. For the time being, then, let's highlight this entrenched, and rather more specific, assumption in imagination theory:

> *The Distinct Attitude Assumption*: Propositional imagination issues in a *sui generis* – i.e. psychologically and ontologically distinct – psychological vehicle, on a par with the vehicles of belief and desire.

Thus, as with belief and desire, is typically characterized in terms of its *sui generis* causal-functional role.

We'll come back in chapters to follow, to look at the motivation for this intuitive assumption. Just setting out and just for now, it will be enough if we isolate and highlight a *prima facie* tension which arises when you couple this assumption with the assumption that imaginings are like slugs:

Question: If we explain cognitive similarities across imagination and belief – as we do with slugs and coins – by positing a shared set of causally relevant properties, then which properties will serve in the account of their definitive causal-functional differences?

The *prima facie* tension embodied here marks the point of departure for most of the arguments to follow in this dissertation.

There may be a temptation to assume that this question embodies a *phantom* tension. Distinguishing psychological kinds by appeal to distinct causal-functional roles is a central tenet of cognitivist explanation. Am I demanding, really, a defense of the foundational assumptions of all of cognitivist psychology? I don't think so, and I will return to address this worry more fully in chapter one. Having said that, I actually think the question does embody a phantom tension. Not because I'm demanding a full and satisfactory account of the metaphysics of classical functional role psychology. Rather, the question is misleading because imagination doesn't have a definitive, *sui generis*, causal-functional role at all.

4. *How Prevalent is the Distinct Attitude Assumption?*

It should be noted that the major strand of recent empirical work in the philosophy of imagination has emphasized similarities across imagination and other attitudes – and

especially belief (Nichols & Stich, 2000, 2003; Nichols, 2004, 2006; Weinberg, 2006). However, rather than pushing toward an analysis of imagination in terms of some other attitude(s), this work has vigorously *upheld* the thrust of *PAC*. Shaun Nichols (2004), for the most salient instance, has argued that similarities exhibited by beliefs and imagination are due primarily to the contingent peculiarities of our cognitive architecture, and not to any deep ontological ties between belief and imagination.[5]

As Nichols acknowledges, his account is an embodiment of a widespread approach to understanding imagination. A breezy tour of the literature is enough to give a sample of the current climate, as it is *rife* with passages which belie that *PAC* is either openly operating, or is not far below the surface. Here is a short list:

Colin McGinn (2004, p. 131) declares, for instance, that:

> '…the verb "imagine" connotes a distinct type of attitude…[D]issimilarities with belief…underline the *sui generis* status of the attitude of imagining-that….[Imagining-that] belongs to another mental category altogether…''

In a similar vein, Gregory Currie and Ian Ravenscroft (2002, p. 17) declare that:

[5] These accounts have generated further, empirical puzzles. It's unclear, in particular, how imagination could be so very like belief in some respects – e.g. when it's rational, action-producing, and affect-producing – and so very unlike belief in others. I develop this problem, in a more empirical mood, in chapter 1.

> ‘[t]he space that beliefs and imaginings both lie in is a space of states which are functional kinds…There may be…kinds of beliefs, with different functional characteristics…and there might be…states of belief-like imagining that correspond to each of these belief states…But they do not intersect.’

Likewise, Jonathan Ichikawa (2009, p. 111) writes:

> ‘As there is a real distinction between images and percepts, so is there likewise a real distinction between beliefs and imaginings. This fact is widely recognized, and I trust it needs no defense here.’

Echoing these observations, Shaun Nichols (2006b, p. 8) – so far the most vociferous proponent of the distinct attitude approach – affirms this increasingly detailed, and accepted orientation to imagination:

> ‘Among cognitive scientists and philosophers of psychology, there is a growing consensus about a basic account of the imagination….Imaginational states are contentful representations, but they are not distinguished from beliefs by their contents. Rather…imaginational representations are distinguished from belief representations by their *functional roles*. Just as desires are distinguished from

beliefs by their pattern of causal interaction, so too imaginings are distinguished from beliefs by their pattern of causal interaction.'

I propose to take Nichols at his word (regarding the scholarly climate), and I hope to have said only enough here to show that recent theories of imagination exhibit a marked tendency to conform to the admittedly natural intuitions behind *PAC* and its supposed corollary in the distinct attitude assumption.

5. *The Game Plan*

So much for the preliminaries. Here is how I propose to proceed from here.

In my first chapter, I present the distinct attitude account of imagination in more detail, arguing that there's a deep explanatory tension in the core commitments of the view.

In chapter 2 I present and criticize the single attitude account of imagination, recently put forward by Peter Langland-Hassan. The single attitude account is an intriguing alternative to the distinct attitude approach, proposing to assimilate the mechanisms of propositional imagination to those of counterfactual reasoning. Ultimately, I argue, the account fails to accommodate large swaths of the data ordinarily associated with imagination.

In chapter 3 I switch gears, in order to look in detail at an historically – and no less presently – important critical case study in the theory of propositional imagination.

Developmental accounts of pretend play in the psychological literature can claim much of the credit for setting the agenda of imagination theory across disciplines into the philosophy of cognitive science. When providing an account of propositional imagination, the tendency has been to start by accommodating the data associated with pretend play in young children. The crux of the issue has ordinarily surrounded whether children require recourse to metacognition, in order to recognize and engage in pretense. Specifically, do they need to utilize a psychologistic understanding of pretense? I take sides with the metarepresentationalists, arguing that no satisfactory, non-metarepresentational account has been yet provided, and that none is likely to be forthcoming.

Finally, in chapter 4 I sketch a pluralistic account of imagination. Drawing on insights from available accounts – insights drawn both from the extant accounts succeed and where they fail – I argue that the primary psychological vehicles of some of the diverse phenomena scholars have tended to lump under the auspices of 'the propositional imagination' decompose into at least three distinct psychological phenomena: counterfactual elaboration, bare imaginings (i.e. mere entertaining that *p*), and fictionalized attitudes akin to the states initially posited by Alan Leslie (1987) in his seminal statement of the metacognitive account of pretense.

Chapter 1: The New Cognitive Theory of Imagination

'[There is] a convergence of opinion that appears, as much as anything ever does, to approach consensus in the current philosophical community….the view has emerged that [imaginative] acts have their power…through their activation of special cognitive attitudes, akin to beliefs in structure and in some of their effects, but distinguished from beliefs in others. This view…[amounts to] positing what we will call a "distinct cognitive attitude"'...(T. Schroeder & C. Matheson, 'Imagination and Emotion,' 2006)

As noted, there's been a relative surge of scholarly interest in the imagination lately. Much of the present discussion is traceable to Alan Leslie's (1987) seminal work on the psychology of pretense. Following Leslie's lead, researchers have since tended to explore links between pretense, the propositional imagination, and folk psychology (Perner, 1988, 1991, 1993; Harris, 1991, 1993, 1994, 1995, 2000; Leslie and Thaiss, 1992; Goldman, 1992; Leslie and Roth, 1993; Harris and Kavanaugh, 1993; Leslie, 1994, 2002; Gordon, 1986; Gordon and Barker, 1994; Currie, 1996; Nichols *et al*, 1996; Nichols and Stich,

2000, 2003; Friedman and Leslie, 2007; Friedman *et al*, 2010; Langland-Hassan, 2011; van Leeuwen, 2009, 2013).

The conversation has proven fruitful, and researchers have reached a surprising degree of consensus on a number of substantive points. First, all parties have tended to begin by adopting a generic form of *cognitivism*, on which pretense and imagination are assumed to be subserved by an internal system of mental representations. Second, all agree that the system trades in representations that (*i*) are defined over propositional contents, (*ii*) are functionally similar to beliefs in important respects, but (*iii*) functionally distinguishable from beliefs in other important respects. Finally, and more contentiously, many have adopted the natural idea that imagination is a distinct propositional attitude, on a par with belief and desire. This idea has issued in a well-articulated, novel empirical paradigm.

I argue here that – for all its explanatory promise – this new empirical paradigm is ill-founded. For, the two core commitments of the view are in tension with one another.

1. The New Consensus – A Distinct Attitude in a Common Code

Since the target account here embodies a widespread family of views, which tend to agree on core assumptions, let's call it the *new cognitive theory of the imagination* (or simply *new cognitivism*, for short). Though new cognitivists differ in the presentation of

their proposals, they converge on two critical empirical assumptions: *the distinct attitude* and *the single code hypotheses.* I'll highlight the substance of these commitments in turn.

Start with the central idea of new cognitivism: the distinct attitude hypothesis. The idea is that propositional imaginings comprise a *sui generis*, functionally individuated psychological kind. New cognitivists are wont to appeal, for instance, to *prima facie* obvious facts about the distinctive causal role of imagination: E.g. that it is often subject to the will, that imaginings are not (typically) taken as input by action production and decision making mechanisms, and that imaginings are largely quarantined from long term memory and from the store of beliefs in general. Since it's the theoretical cornerstone of new cognitivism, let me make the distinct attitude hypothesis explicit:

> *The Distinct Attitude Hypothesis*: Token states of the propositional imagination are intentional psychological states that are functionally individuated *vis-à-vis* the other paradigmatic intentional psychological kinds – e.g. beliefs and desires.

This core assumption – at least as it is ordinarily developed – is intimately related to the other core commitment of the view: It's typically assumed that imaginings are *semantically equivalent* to other propositional attitudes, and especially to beliefs. That is, it's usually supposed that when, for example, one imagines that *Lisbon is in Ohio*, one is in a psychological state with the very same content as the belief (or desire, or hope, or

fear, etc.) that *Lisbon is in Ohio*. There's an obvious reason for this. If token imaginings exhibited some systematic semantic difference from beliefs, and, hence, were partially constituted by some special concept(s) or syntactic operator(s) – e.g. PRETEND, IN THE FICTION, or IMAGINE – then the explanatory demand for augmenting the inventory of psychological states would be undermined. One could instead articulate the distinction between propositional imaginings and other mental states in terms of their proprietary contents (Leslie, 1987; Freidman *et al*, 2010).

The distinct attitude hypothesis seems plausible for a variety of empirical reasons, with one intriguing line of argument focusing on the synchronous processing of belief and imagination (McCune-Nicolich, 1981; Leslie, 1987; Nichols and Stich, 2003; Nichols, 2004; Nichols, 2006a; Weinberg and Meskin, 2006a, 2006b). It's noteworthy, simply put, that we can apparently imagine that *p* and believe that *p* at the same time. Playing a game of tea parties, I may imagine that I spilled my tea, and that my cup is thus empty. But I also *believe* that the cup is empty. In which case, it appears that I harbor numerically distinct tokens of semantically equivalent representations. Assuming so, we'd have to look to non-semantic differences in order to individuate the relevant representations, and functional differences provide an obvious option. And the phenomenology of imagining that *p* seems to corroborate. Ordinarily there is, as it were, no appearance (while imagining that *p*) of imagination-related concepts.

The distinct attitude hypothesis, together with the semantic equivalence assumption, lead to an intriguing idea: Perhaps *all* that distinguishes imaginings from

beliefs (and desires, for that matter) is their causal/functional profile. That is, in addition to their contents, perhaps imaginings share other properties with beliefs, and in particular maybe they share properties which explain why imaginings are actually *commensurate* with belief with respect to a suite of cognitive processes. We might call these other properties 'logical' or 'syntactic', but to leave open the possibility that mental representations are non-linguistic in nature (Nichols and Stich, 2003, p. 32), I'll lump the relevant properties under the more general headings of 'structural' and/or 'compositional'.

Notice the crucial, corollary assumption that accompanies the assumption of compositional equivalence: What we're calling structural or compositional properties are ordinarily assumed to bear some explanatory relation to the causal/functional properties of the representations in which they inhere. Together, the compositional and causal properties should play, especially, in the account of the representation's interactions with specific cognitive mechanisms – e.g. in the case of the inferential mechanisms, the causal effects produced ought to honor compositional properties of the representations borne by the vehicles. Thus, by assuming that imaginings and beliefs are structurally or compositionally isomorphic, we are assuming that they share properties relevant to the determination of both their contents and their causal profile.

This assumption is embodied in the second foundational commitment of new cognitivism, *the single code hypothesis*:

> *The Single Code Hypothesis*: Token imaginings that *p* are compositionally isomorphic to token beliefs that *p*, and tokens of each kind are processed in much the same way by those cognitive mechanisms which receive inputs of both kinds.

Positing a distinct attitude in a common code has been heralded as an important development in imagination theory – and for good reason. The assumptions of new cognitivism deliver a powerful explanatory package. They help to explain, for one thing, why imagining can have strong emotional effects. It's a curious fact about us that we respond emotionally to states of affairs we know to be false or fictional, and we frequently treat characters we know to be non-existent as though they were real. According to new cognitivism, this is because affect mechanisms don't distinguish compositionally isomorphic beliefs and imaginings. If – reading *Othello* – I'm disgusted by Iago's behavior, it's because affect production mechanisms don't track the fact that I'm (merely) imagining the contents of a fiction.[6] Similarly, if I judge Iago to be contemptible and blameworthy, it's because the mechanisms of moral cognition don't distinguish what I believe from what I merely imagine, and so on.

Much has been made, too, of the *inferential* symmetries across belief and imagination. Imaginers spontaneously draw on memory and background beliefs in order to elaborate the contents of imagination (Leslie, 1987; Harris, 1991, 2000; Nichols and Stich, 2000, 2003). The emerging pattern of explanation is clear – the imagination

[6] See Walton, 1990 and Walton, 1997 for notable exceptions to the assumption that such processes involve the production of genuine emotions.

exhibits seamless liaisons with the store of beliefs because (*ex hypothesi*) the inference mechanisms treat both kinds of representation similarly. New cognitivism thus promises to help account for some of the curious facts about the relation between belief and imagination, and between imagination and our emotional and moral lives. With so much to offer, it's no wonder that new cognitivism can claim to be the default account. But the consensus and optimism have been premature. Or so I'll now argue.

2. *The Asymmetry Challenge*

We've seen, in rough outline, that the new cognitive theory has excellent resources for explaining the functional similarities between belief and imagination. This is part of what makes it an attractive accompaniment to a cognitive theory of pretense; for, it's long been recognized that even very young pretenders draw on their background assumptions about reality in order to make inferences about the contents of a pretend scenario.

But to stop here would be to provide an unsatisfactory account. If imaginings truly compose a distinct psychological attitude, we also need some satisfying account of what's distinctive about imagining that *p*. Imaginings ought to be, as it were, as distinctive from belief as belief is from desire. *Prima facie* it's unclear whether this is the case – as noted, (and unlike desire) there are lots of cases when imagination seems *very* belief-like. Showing that, and how, imagination comes apart from belief will involve, in

particular, detailing the imagination asymmetries, and explaining the mechanisms behind them.

For its part, the distinct attitude hypothesis is a general assumption about the functional status of imaginings, and thus which – taken alone – provides no resources for the sort of account required (more of which below). The single code hypothesis, then, is intended to explain symmetrical processing. This leaves us, presently, without a sufficiently detailed explanation of the asymmetries upon which the viability of the distinct attitude hypothesis depends.

Before proceeding, here's, to a first approximation, the nature of the challenge that I have in mind:

> *The Asymmetry Challenge*: Assume that the single code hypothesis is correct: how do we explain the occurrence of the functional asymmetries which the distinct attitude hypothesis demands? Specifically: (*i*) Why do some cognitive mechanisms consume and produce either beliefs or imaginings exclusively? (*ii*) Why do those cognitive mechanisms that consume both beliefs and imaginings ever treat them differently?

One caveat: It bears emphasizing that in order for the new cognitive theory to gain any traction at all, there must be *some* functional asymmetries to characterize and then

theorize around. Fortunately, there are some clear *prima facie* candidates. I should stress, however, that I wish to remain neutral on whether the candidates that follow amount to *genuine* and *robust* functional asymmetries – by which I mean functional asymmetries of an order suitable to ground the claim that imagination is a distinct cognitive attitude.

Now, to get an idea of the scope of the challenge, it will be useful to provide a putative list of which asymmetries need to be accounted for – in no special order – here are several to consider.

2.1 Intention Direction

In contrast with beliefs, imaginings seem to be producible at will. In response to an intention to do so I can readily imagine, for example, that *there is a giraffe in the parlor*. This imagining might serve as a guide to a whole host of other voluntary or spontaneous imaginings – e.g. that *the giraffe is my pet*, that *I'm a giraffe*, and so on. Fortunately, I cannot so easily produce in myself the *belief* that *I'm a giraffe*. Thus, the imagination is apparently connected to a representation producing mechanism which is subject to intention direction. Following Nichols and Stich, 2003, and Weinberg and Meskin, 2006a, let's call this mechanism the *Inputter-Elaborator*.[7]

[7] I leave cases of imaginative resistance unaddressed. In such cases, imagining that *p* is difficult, and intentions to imagine that *p* may be frustrated. See Moran, 1994; Gendler, 2000; Weatherson, 2004; Walton, 2006; Weinberg and Meskin, 2006a, 2006b. Suffice it to say that these cases compose a special class.

An obvious question arises: why are imaginings, but not beliefs, subject to intention direction? Take, for instance, the (voluntarily produced) imagining that *p* and the (involuntarily produced) belief that *p*. There are, by hypothesis, no semantic or structural properties of either representation that serve to indicate that the one could be produced by an intention-sensitive mechanism while the other could not. So why should imaginings be subject to the will at all, when isomorphic beliefs are not? After all, we can form intentions to believe that *p* as easily as we can form intentions to imagine that *p*. It would be, perhaps, too much to shoulder new cognitivism with the burden of accounting for the relation between imagination and volition. Nevertheless, the intention-directedness of imagining presents an instance of the asymmetry challenge, and for all that's been said so far about new cognitivism there's little hint about how the asymmetry should be explained.

2.2 Action Production

Beliefs and imaginings apparently relate differently to action. To be sure, imaginings are capable of guiding behavior (as in pretense), but imaginings clearly do not bear the same kind of causal relations to action and decision-making that beliefs do. If I'm imagining within the context of a pretense that *there is a fire in the chimney*, I may feign dialing the fire department. But I won't really make the call. No decision to call is acted upon, presumably, because the relevant imaginings lack some causal properties which the corresponding beliefs possess. But which ones?

A tempting answer is that the question, 'Which ones (which causal properties)?' is confused. After all, we're *talking about causal properties*, and thus perhaps the right thing to say is just that the action production mechanisms don't receive inputs from imagination – and that *just is* the causal property that beliefs have/imaginings lack. I'll have more to say about this sentiment. For the time being, notice that there is no obvious reason to suppose that the action production mechanisms could not take both imaginings and beliefs as inputs. After all, by hypothesis, they share many other cognitive pathways, are treated similarly by other cognitive mechanisms, and so on. New cognitivists have had little to say about how this should be explained. That is, about why it is not simply arbitrary to assume that two putative psychological kinds which are so causally, semantically, and structurally similar bear so very different relations to volition.

2.3 Attitude Identification

We have the capacity to introspect and reason about the contents and functional status of some of our intentional mental states. When we introspect beliefs, for instance, we generate beliefs about what we believe. Similarly, when we introspect imaginings, we generate beliefs about what we imagine. Thus, we apparently have knowledge – to some degree and of some sort – of our beliefs and imaginings *qua* beliefs and imaginings. For example, a token representation with the content *that it is snowing on Denali* might occur as either a belief or an imagining, and we can usually tell the difference. Presumably,

there is some cognitive mechanism – call it the *Monitor* – the operations of which help to explain the acquisition and storage of this sort of knowledge (Nichols & Stich, 2003).

But when I imagine that *it is snowing on Denali*, and I introspect this state, what explains why the Monitor generates the belief that *I imagine that it is snowing on Denali* instead of the belief that *I believe that it is snowing on Denali*? For that matter, what determines that the second-order representation generated by the Monitor is a second-order belief, but not a second-order imagining? To be sure, the imagination theorist isn't required to develop a complementary theory of attitude identification. However, it's to be hoped that one's imagination theory will not prove resistant to combination with accounts of related aspects of the mind. Unfortunately, few hints have been provided by new cognitivists as to how attitude monitoring might be so reliably achieved in a single code architecture.

2.4 Affect

As we saw above, belief and imagination can have similar affective consequences. But this is not always the case. In contrast with beliefs, imaginings may exhibit *curtailed* and *diluted* affective consequences – frequently, the consequences of imagining that *p* are neither as enduring nor as intense as those of belief. As Nichols (2006a) points out, these asymmetries are not terribly troubling for new cognitivism. Since according to everyone the imagination is (typically) subject to the will, curtailed affect is handily explained by our capacity to disengage the imagination at will. Diluted affective consequences are also

predicted by new cognitivism, since in general we tend to identify less personally and less richly with the objects, agents, and states of affairs in imagination.

There are, however, two affective asymmetries that have received special attention, since they require *prima facie* more worrisome qualifications. Nichols (2006a) observes that imaginings sometimes exhibit *discrepant* affect: there are cases in which affective responses to imagining that *p* are robust, but strikingly different from the predictable affective responses to the belief that *p*. For instance, if one is imaginatively engaged in a black comedy, one might be amused by its dark contents. Engaging a tragedy may engender fulfillment at the witnessing of the lamentable outcome. In cases of *absent* affect, imagining that *p* has *no* affective consequences, while believing that *p* would likely entrain a considerable response. Imagining that *p* is likely to exhibit absent affect, for example, when imaginatively drawing out the consequences of a philosophical trolley car experiment.

Why should affect production systems produce differential outputs to structurally and semantically isomorphic beliefs and imaginings in these ways? After all, part of the attraction of the single code hypothesis is that it explains why the affect production mechanisms sometimes treat isomorphic beliefs and imaginings similarly. In contrast with other belief/imagination asymmetries, affect *has* attracted the attention of new cognitivists (Nichols 2006a), and I'll have occasion to return to consider the response more fully below.

2.5 Inference

The spontaneous revision of the contents of imagination mirrors that of belief revision. Both are orderly and predictable, exhibiting seamless access to background knowledge, and so on. *Prima facie*, this is best explained by assuming that a single inferential mechanism – call it the *Updater* – takes input from both imagination and belief, and treats them similarly.

But there is a substantive asymmetry to consider. While the Updater readily revises what we imagine in light of what we believe, it typically does not update our beliefs in light of what we imagine. For instance, when we imagine that *the cup is full of tea*, and we see our playmate upend the cup, we are not thereby led to *believe* that *there is tea on the table*. And from an epistemological standpoint, we don't consider the upending to comprise additional evidence for our prior belief that the cup is empty – we held that belief based on independent perceptual grounds. As researchers are wont to say, revision in imagination occurs under *quarantine*, within which representations are automatically channeled to imagination. Or, if you prefer a more computational metaphor, revisions in imagination are automatically *written* to the imagination box. There are no doubt exceptions, and these will play a larger role in the discussion below. Nevertheless this does constitute a generally accepted guideline for imaginational updating (Lewis, 1983; Currie and Ravenscroft, 2002; Nichols, 2004).

So why does the Updater effect a proprietary procedure when imaginings feature among its inputs? If parallel imaginings and beliefs are semantically and structurally

equivalent, how could the relevant mechanism be sensitive to the functional status of imaginings? Once again, for all new cognitivism has to offer, there've been few suggestions about how to accommodate the asymmetrical processing by the Updater.

2.6 Taking Stock

My aim so far has been to make vivid the nature of the asymmetry challenge. Assuming that beliefs and imaginings occur in a single code, we should certainly predict functional similarities. But if the single code hypothesis explains commensurate processing – and assuming that there are robust functional differences across imagination and belief – what's left to explain asymmetrical processing? In effect, what aspect(s) of the belief-imagination system should lead us to predict that beliefs and imaginings will be treated differently in the relevant cases? To answer these questions would be to answer the asymmetry challenge.

The problem, however, is that no reply seems consonant with the letter of new cognitivism. Without pretending to exhaust the possibilities, I'll evaluate what I take to be some plausible replies to the challenge. I separate these replies here, but there's no reason to assume they couldn't be combined in various ways. Ultimately, however, I take it that there's no satisfactory combination available to the proponent of the distinct attitude hypothesis.

3. The Simple Reply

The simple reply is likely to be the default new cognitivist response. It begins with a natural thought, formulated articulated by Nichols (2004, p. 131) as follows:

> '…the [single code hypothesis] can only be framed against a background of cognitive architecture. Once one has posited a background of cognitive components, *one can then explain some processing differences between imagination and belief by noting that some of these cognitive components take input from beliefs but not from imagination*.'[my emphasis]

The general idea behind the simple reply is thus to rely on direct appeals to contrasting input-output relations among the mechanisms comprising the belief-imagination system.

There are a number of points to make about this proposal. So far as it goes, the proposal promises to neatly explain at least one sort of asymmetry: Namely the sort in which a cognitive mechanism apparently takes either beliefs or imaginings as input, but not both. We might call these asymmetries of consumption – i.e. asymmetries due to a lack of causal access to an input pathway. One might try to account for the asymmetry in action production, for instance, by assuming that beliefs – but not imaginings – are consumed as inputs by the relevant mechanisms.

But the elegance of the explanation is misleading, and a closer look shows that this direction of explanation is unsatisfactory. What we really need to know is *why* imaginings aren't, or cannot be, forwarded – or written – to the action production mechanisms. Since imaginings and beliefs are supposed to share so many other causal pathways, it's simply not enough to say that some pathways are blocked. What we need is an account of the mechanism by which the relevant rule is effected. For instance, in this case, the rule *that beliefs, but not their compositional counterparts in imagination, are consumable by the relevant decision making mechanisms*. Simply to say that these mechanisms eschew imaginings is only to label the asymmetry in processing.

Moreover, the simple reply leaves two important kinds of case unaddressed: (*i*) *production* asymmetries, where a mechanism exclusively *produces* either beliefs or imaginings, and (*ii*) *intra-mechanism* asymmetries, where both beliefs and imaginings are consumed as inputs, but are treated differentially by a mechanism. First, I'll consider whether production asymmetries can be accommodated within the present framework.

Take the mechanisms of intention-direction for imagination as an example. Perhaps we should posit an exclusive *output* pathway from the Inputter-Elaborator to the relevant functionally individuated workspace. This strategy for explaining production asymmetries has the same *prima facie* appeal as the case of action production just rehearsed. And it raises a similar worry. What we're after is an account of the mechanism by which the asymmetry occurs. But so far we have little more than an acknowledgement

that it occurs. At best, then, what we've got so far is a promissory note that the asymmetry challenge is answerable within the distinct attitude-single code frame work.

Also, there is a more worrisome class of case to consider. Setting the consumption and production asymmetries like those described aside, it's far from clear how to accommodate *intra-mechanistic* asymmetries. Consider the Updater, for example, which is supposed to take beliefs and imaginings as input, to treat them similarly for some purposes – e.g. as in spontaneous revision – yet to treat them differently for others – e.g. keeping a proprietary output pathway for imaginings. One thing we of course cannot assume is that there's no output pathway from the Updater to the belief box. Another strategy would be to propose that, as a rule, the Updater closes the path to the pretense box whenever it receives input from imagination. Indeed, since this seems to be something approaching the default assumption (Currie & Ravenscroft, 2002; Nichols, 2004, Carruthers 2006, van Leeuwen, 2013), let's call this idea the *pretense priority hypothesis*.

The pretense priority hypothesis brings the limits of the simple reply into focus. As above, in spite of its *prima facie* plausibility and elegance, what we've really been delivered is a mere description of the phenomenon. It seems like the blockage the hypothesis describes typically occurs (or at least it seems like it does). But we have no idea how to describe the mechanism by which it occurs, if and when it does. Hence, we don't understand how the asymmetry is realized.

Why demand more than the assumption that the Updater reacts differentially to imaginings? According to the single code hypothesis, mechanisms which take both beliefs and imaginings as inputs will (typically) treat them similarly. The Updater violates this assumption to the extent that upholding the distinct attitude hypothesis demands that it effects a proprietary procedure in response to imaginings. Sets of inputs consisting only of beliefs contribute to the production of further beliefs; sets of inputs including imaginings contribute the production of further imaginings. The pretense priority hypothesis, at first blush, embodies a violation of the single code hypothesis. Thus, we surely need something more than just the assumption of differential treatment if we're going to uphold the consistency of new cognitivism.

There's surely some room to maneuver on this point, though it's unclear how much room. A tempting idea is that imaginings are somehow labeled, and that the labeling explains pretense priority, and perhaps other asymmetries of intra-mechanistic processing. But given that they're supposed to be compositionally equivalent – i.e. equivalent in semantics and structure – there doesn't seem to be any properties of the vehicles of the contents of beliefs or imaginings that could serve to alert the Updater to the presence of imaginings, thus triggering the proprietary inferential procedures. Nevertheless, the idea of imagination labels merits some exploring, since we seem to have pushed the simple reply to its limits.

4. The Labeled Origins Reply

Let's try elaborating the simple reply with the proposal which allows the shared mechanisms comprising the belief-imagination system to track the functional status of (at least) imaginings. The idea has the natural corollary mentioned above: Perhaps the system incorporates a mechanism, call it the *Marker*, which annexes a *label* to imaginings for the purposes of downstream processing.[8] If a mechanism like the Updater was sensitive to the imagination label, then we might have a better foundation for explaining its capacity for differential intra-mechanistic processing. Moreover, the imagination label might help us understand other kinds of asymmetry. Some mechanisms – e.g. the Inputter-Elaborator or the Monitor – may only produce representations with the relevant tag. Other mechanisms – e.g. action production mechanisms – may reject representations with the relevant tag.

This strategy strikes me as promising, but the question is whether it's consonant with the fundamentals of new cognitivism. Let's consider two ways of understanding how the Marker might work. For one, it may brand token imaginings with (as it were) a *thick* label. We'll say a label is thick if it alters the compositional structure of a mental representation. Or, the Marker might imprint a *thin* label. We'll say a label is thin if it has

[8] Leslie (1987) endorses a similar proposal, hypothesizing that a mechanism (the *Expression Raiser*) furnishes pretense representations with quotation markers which decouple them from the ordinary input/output pathways for belief.

no real structural or semantic import, though it nevertheless is something to which a cognitive mechanism could be sensitive.

Thick labels seem obviously ruled out if we're operating within the framework of new cognitivism. In addition to being inconsistent with the letter of the single code hypothesis, they would undermine its primary virtue. Positing a strict semantic and structural isomorphism across belief and imagination is supposed to explain why their effects are similar in many respects. Denying the isomorphism by adding thick labels to the architecture would thus resuscitate the challenge the single code hypothesis was designed to answer.

Moreover, thick labels would put the distinct attitude hypothesis in jeopardy. If imaginings have proprietary compositional properties, and in particular proprietary *contents*, a central motivation for individuating them functionally would be undermined. Endorsing thick labels would, in effect, recast imaginings as beliefs with proprietary semantic and/or structural properties. At the very least, since special representational properties would be carrying the large share of the explanatory burden, the explanatory role of the distinct attitude hypothesis would be much less clear. The point to take away is *not* that the challenge of explaining functional similarities and differences across belief and imagination is insurmountable if we adopt thick labels – actually this is just the strategy I mean to support (see also Friedman and Leslie, 2007; Friedman *et al*, 2010; Langland-Hassan, 2011). The point is just that any theory which takes on thick labels wouldn't be new cognitivism.

So much – for the time being – for thick imagination labels. What about thin labels? Perhaps we can put a broader sense of 'structure' to work, a sense which encompasses more than just those structural properties related to the *compositional* structure of imaginings. Thin labels, we might say, serve to mark functional status by altering the structure of imaginings, but without altering their compositional properties. Adding thin labels to the architecture has some *prima facie* plausibility.

Start with a question: What will the account of the *Marker* look like? A natural proposal is that imaginings are annexed with their labels upon production. As the Inputter-Elaborator writes an imagining to the pretense box, the label is synchronously affixed. Contrast, for instance, the vehicle for the belief that *it's snowing on Denali*, with the vehicle for the compositionally isomorphic imagining**.** By hypothesis, let's say that the Inputter-Elaborator affixes an imagination tag – which I'll signify by '*****' – to the latter upon its production. We'd thus have the compositionally equivalent, but nevertheless distinguishable, representational vehicles IT'S SNOWING ON DENALI and IT'S SNOWING ON DENALI*****.

So far, so good. But difficulties arise on both metaphysical and mechanistic fronts here. Let's begin with a look at the mechanistic worries.

First consider that not all imaginings are produced by the Inputter-Elaborator. We know that the Updater spontaneously produces imaginings in a process resembling belief revision. I've already argued that what we need to accommodate this process is a story about how the Updater reliably writes its productions to the pretense box in such cases. If

we're following the present strategy, this story presumably must link the Updater and the Marker, or else it must provide the Updater with its own labeling mechanism. For, the imaginings produced by the Updater must be labeled in the same way as those produced by the Inputter-Elaborator. In which case, thin labels haven't helped us out as much as one might have thought. For, we now need an account of the relation between the Updater and Marker, *including* an account of how the imaginings produced by the Updater are reliably distinguished by the marking mechanism for labeling. In other words, incorporating thin labels yields a new asymmetry – one which very closely parallels the asymmetry we set out to explain. Similar considerations will apply, *mutatis mutandis*, to any cases of intra-mechanistic processing resulting in the production of further imaginings.

And there's an additional worry lurking here. Labeling proposals, generally speaking, are all the more attractive assuming a robust suite of functional differences for the labeled representations. That is, incorporating labels seems well-motivated when the relevant mechanisms reliably and predictably effect proprietary procedures when operating upon the labeled representations (and, as in the case of the Updater, the representations *to be* labeled). The lurking worry arises in the event that the asymmetries which thin labels have been imported to help explain turn out to be not so very robust after all. If the asymmetries are insufficiently robust, then the inclusion of thin labels may be *ad hoc* or just superfluous.

This is a live possibility, at least with respect to the workings of the Updater. Contrary to the pretense priority hypothesis, for example, there are many cases in which the Updater takes input from imagination, but produces *beliefs* (see also van Leeuwen, 2013). Imaginings feature, for instance, in inferences yielding modal beliefs. I may come to believe *that talking pigs are possible*, after imagining a world in which pigs converse with their human caretakers. Everyone, of course, is aware of this exception to pretense priority. But there are other exceptions – and enough, I take it, to draw pretense priority into question. In addition to modal inferences, imaginings feature prominently in the causal-epistemological base for counterfactual inference. Imagining, for instance, *that it snows at least six inches by dawn*, may play in my judging that *if it were to snow at least six inches by dawn, then I'd be late for work*. And beliefs representing conditional probabilities (beliefs of the form *if p, then probably or probably not* $q_1,...q_n$) are frequently produced in a process that involves representing *p – as something intuitively akin to an act of imagination* – before assessing the likelihood of various outcomes.[9]

You might think all this is fine. Maybe modal, counterfactual, and probabilistic reasoning comprise special (and perhaps related) cases in which the inference mechanisms reliably disregard the (thin) label annexed to the vehicles of imagination. While fully addressing the issues involved here would take us too far out of focus, it's worth noticing that the causal scope of imagination likely outstrips the cases just rehearsed. In particular, there's no reason to assume that imagining that *p* doesn't

[9] For various proposals, see Rips and Marcus (1977), Kahneman and Tversky (1982), and Johnson-Laird and Byrne (1991, 2002). Evans and Over (2004) provide a critical review of the psychology and philosophy of hypothetical reasoning.

sometimes play in the production of ordinary beliefs (by which I mean beliefs that *p*). Suppose I imagine that, on Twin Earth, all the stuff they call 'water' (and which we call 'twater') is XYZ. I may be thereby lead to judge that the meaning of natural kind terms depends on facts external to speakers. At the very least, *prima facie* there's nothing *pathological* about this sort of inference. We may be wary of the legitimacy of such inferences, but the question before us is about psychological processing, and not about the normative status of any particular process.

And the same goes for some rather more mundane sorts of case. Suppose I imagine that *my wife wears the yellow dress*, and suppose this is accompanied by my imagining that she's feeling happy and confident, and so on. The cascade of representations in this case may lead to my *concluding that the yellow dress really is her favorite one*. Nothing in this kind of case seems especially extraordinary, and neither does it seem to require considering possibilities (explicitly) or hypothesizing. It's an open question, of course, whether I'd come to the relevant conclusion without the aid of mental imagery. But this is an open question about the relation of the propositional imagination to mental imagery, and any temptation to say that it's *merely* a case of visual imagination surely ought to be avoided. Thus, the robustness of (at least) pretense priority in inference is clearly questionable. In which case, one may well begin to wonder about the robustness of the other putative functional asymmetries, too. It may be, that upon further inspection, the suite of functional asymmetries I detailed above do not describe (really) distinct causal profiles for belief and imagination.

It should be noted, before moving on, that in their seminal presentation of the distinct attitude hypothesis, Nichols and Stich (2003, p. 49) seem happy to embrace a kind of thin labelling proposal for the vehicles of imagination. Moreover, since Leslie's metacognitive account also incorporates labels, they suggest that their primary competition – the Leslie-style metacognitivism about pretense – is in crucial respects just a notational variant of their own account:

> 'Leslie's hypothesis that…the [imaginational] representations subserving pretense are 'quarantined' or 'marked off' is equivalent to claiming, as we do in our theory, that pretense-subserving representations are [compose a distinct attitude] in a box of their own…The representations in the Possible World Box (in our theory), or within the quotation marks (in Leslie's theory) are tokens of the same types as the representations in the Belief Box (to use our preferred jargon) or in the pretender's primary representations (to use Leslie's)… Nonetheless, the part of Leslie's theory that we have set out so far can plausibly be viewed as simply a notational variant of part of our theory.'

The assumption here seems to be that, if Leslie's imaginational representations really just require a thin label in order to explain their functional distinctiveness, then there really may not be very much of a disagreement between the metacognitive and new cognitive

accounts of imagination (that is, so long as we're setting accounts of how young children recognize pretense behavior to one side – I'll return to this subject in chapter 3).

Now, one might read 'marked off' and 'quarantined' as loose talk with roughly the broad sense of 'functionally distinguished' or 'thinly labelled' that Nichols and Stich (2003, p. 58) apparently have in mind:

> 'Using a marker to indicate that pretense representations have different functional roles from beliefs is, as we noted earlier, the equivalent of positing a separate box for pretense representations, and that suffices to quarantine pretenses from beliefs. Positing an additional difference at the level of content does no additional work…'

But Nichols and Stich seem, on the one hand, to overestimate the role of boxological explications of cognitive architectures. It is really not for the cognitive architecture, *per se*, to determine whether a particular component of a representation bears a content (though, to be fair, the targeted theorists seem to make a similar faulty assumption).

On the other hand, perhaps more importantly, we really have no good grasp over whether the notion of a *semantically-empty* – i.e. thin – label is legitimate, to begin with. Given the functional considerations, Nichols and Stich (2003, p. 58) assert that they 'see no reason to adopt…[the]…view that there is a systematic difference in the contents of

the representations underlying pretense and belief…' But if we're adding representation labels to the account, as they're apparently willing to accept, there's really no obvious reason to assume *otherwise*, either – i.e. to adopt the view that belief and pretense *do not* exhibit different contents. If we want to know whether representations such as IT'S SNOWING ON DENALI and IT'S SNOWING ON DENALI* have systematically similar or different contents, we need to move away from cognitive architecture, and look to our theory of the metaphysics of content.

Maybe we should take a quick look at the options. Suppose the compositional features of mental representations gain their contents *via* a causal covariance relation. Perhaps we should say that the label '*' causally co-varies with the imaginings to which it is annexed. If so, perhaps we should say that it bears content much like the concept IMAGINE or PRETEND bears. Suppose the compositional features of mental representations gain their content *via* their inferential role. If so, perhaps in that case we should likewise say that the label '*' bears contents similar to the concepts IMAGINE or PRETEND. After all, part of the explanatory role of the '*' marker (or Leslie's quotation marks, or the pretense box) is to play a particular inferential role – the role associated with imagination. Thus, to a first approximation, deferring to theories of content may turn out to lend some *prima facie* support to metacognitive accounts of imagination.

But anyway, the point here is not to adjudicate the issue. It is rather to point out that while it may very well be that an important piece of the metacognitive account is a notational variant of the functionally distinct attitude view, this fact may nevertheless

turn out for the worse for the new cognitive theory. It depends on presently unanswerable questions of the metaphysics of content

There's more to be said, to be sure, but I won't explore the labeling option any further. What's clear is that the proposal carries substantial explanatory burdens of its own. Thick labels are out, and thin labels might explain some asymmetries but not others. Moreover, it's unclear whether thin labels are explanatorily legitimate, since we'll apparently be forced to cherry-pick the cases for which we'll invoke them. In effect, the temptation will be to appeal to thin labels only where they seem to serve a purpose – e.g. in cases where we see pretense priority – but not in other cases – e.g. in cases like modal and counterfactual inference where the Updater seems to ignore the label. A brief review of these challenges has been enough to cast doubt on whether they can be accommodated within the distinct attitude/single code framework provided by new cognitivism.

5. *The Desire Effects Reply*

A rather different kind of strategy has already appeared in the new cognitivist literature. Nichols (2006a) appeals to links between imagination and *desire* in order to explain cases of asymmetrical affective consequences across belief and imagination. Notice that imagining that *p* and believing that *p* may yield strikingly different emotional outputs. We may take pleasure in imagining the dark contents of a tragedy or black comedy, for

instance, although we'd likely experience very different emotions if we believed the contents were really occurring. And sometimes, imagining that *p* simply fails to produce the affective consequences we'd expect from parallel belief(s). Paradigmatic cases are manifest wherever we engage in the cold contemplation of the harrowing contents of a philosophical thought experiment (pick your favorite extravagantly tragic trolley case, for example). In such cases, imagining doesn't produce much emotion all, though believing the contents surely would.

To explain cases of discrepant and absent affect, Nichols (2006a, p. 472) suggests that:

> '[t]he explanation for the asymmetries 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.*'(original emphases)

Moreover, the differential inputs sent to the affective mechanism are due to the different sets of desires associated with the content *p*, depending upon whether *p* is imagined or believed. Thus, the general idea seems to be to keep with the letter of the single code hypothesis – eschewing even thin labels – by appealing to proprietary links between particular sets of imaginings and the particular sets of desires which accompany them.

How, more specifically, do desires help to explain the asymmetries? A general overview will have to suffice for our purposes. Start with the observation that desires influence belief revision in the Updater. There's a large body of empirical work which confirms that our cares, worries, goals, and motives structure inferences and memory activation patterns.[10] Since we know that belief revision and imaginational updating are intimately linked, it's plausible that desires have will similar effects on imagination. After all, we certainly have cares, worries, and desires about the possible worlds conjured in imagination – including complex desires regarding conformity of fictional worlds to genres (Nichols, 2006a).

These observations about the effects of desire on the Updater set the stage for the central explanatory moves. On the one hand, since desires influence *what* contents get written to the belief box and the pretense box, *a fortiori* desires will influence what contents get forwarded to the affect mechanisms. The nature and richness of the contents taken up by affective mechanisms will thus differ as a function of the different desires associated with them. In cases of absent affect in response to imagination, we may have few and/or weak desires directed upon an imaginary scenario. I may be unmoved by the harrowing contents of a tragic trolley case, for example, simply because it describes a fictional world I happen to care very little about. I'm only imagining it, after all, because some philosopher thinks it will help to make an argumentative point. In such cases, it shouldn't be surprising that the sparse inferential output generated, together with the

[10] Nichols, 2006a provides a brief but suggestive summary of the relevant literature. See Kunda, 1999, for a more comprehensive review. See also Sanitosa, Kunda, and Fong, 1990; Noordman *et al*., 1992; Ericsson and Simon, 1993; Doosje *et al*., 1995; Zwaan *et al*., 1995; Goldman, *et al*., 1999; van den Broek *et al*., 2001.

sparse and impersonal inputs from desire, should fail to entrain any substantive emotional response. Moreover, if we start enriching the contents, let's say, by adding family members, spouses, or puppies to the scenario, it's reasonable to expect an elevated emotional output. As for discrepant affect, such cases will also be due to the variant sets of desires we have regarding a particular imaginary world and the actual world. While I may want Desdemona to live, for instance, I may have an overriding desire that the fiction conform to its tragic genre. When it does conform, I take pleasure in spite of myself. If I believed Desdemona to be a woman in the actual world, I would have very different desires regarding how things turn out for her. In general, I certainly don't want actual states of affairs and events to conform to the norms governing the narrative trajectories of Shakespearean tragedies.

I don't propose to challenge Nichols' strategy for explaining the particular cases it's designed to accommodate, though this isn't to say that I take the account to be sound. There may be an issue here concerning how it is that imagining that *p* and believing that *p* come to be tied to different sets of desires in the first place. If so, the desire effects strategy answers one instance of the asymmetry challenge without addressing a prior instance. But instead of challenging the details of the desire effects proposal, I wish to make two, rather more general, points. First, the desire effects strategy fails as a *general solution* to the present challenge. Second, taken literally, the proposal puts the distinct attitude hypothesis in jeopardy.

As rehearsed, the core idea is to explain some variant causal properties of beliefs and imaginings by appeal to the *contents* of variant sets of desires. Whether the proposal

is a candidate for a general answer to the asymmetry challenge will thus depend on whether variant desire contents can explain other asymmetrical effects of imagining and believing.[11] But in many cases the proposal loses much of its *prima facie* plausibility. It would be odd, for instance, if desires were to explain how the Monitor distinguishes beliefs and imaginings for introspection. One should hope, in particular, that whether the Monitor identifies a belief *as* a belief or an imagining *as* an imagining has little to do with our cares, worries, and motivations.

Similar worries apply to the Updater. Granting that desires influence the direction of belief revision and imaginational updating does nothing, *per se*, to explain how the Updater reliably writes variously to the belief box or the pretense box (as appropriate). Which desires might be adequate to play that role? Not desires to imagine specific contents – e.g. the desire *that I imagine that p*. That proposal predicts that either I'm not likely to infer – or perhaps even that I cannot infer – some specific imagining without the accompanying desire to imagine that *p*. This is a bad prediction. For, imagining that *p* is not made any less probable where the subject lacks the associated desire. We can imagine that *p* even while desiring, to the contrary, not to imagine that *p* (or to imagine that ~*p*).

Another idea would be that the Updater is influenced to write to the pretense box whenever we harbor a general desire to imagine (but not necessarily to imagine anything in particular). But the objection just rehearsed generalizes. *Prima facie*, imagining doesn't require a desire to engage in the activity of imagining at all. I may find myself unexpectedly consumed with the contents of a made-for-TV movie, or a short story, for

[11] And it bears emphasizing that this *not* a proposal that Nichols explores himself.

instance, despite the fact that I have no standing desire to imagine anything. My engagement may continue, indeed, even in the face of the desire *that I desist in the activity of imagining*.[12] These are surely not insurmountable counterexamples to the proposal. The point is just that some rather ordinary cases highlight the tenuous relation between desire and the mechanisms responsible for the spontaneous production of imaginings. Desires may well influence the *contents* of imagination, but this does little to help explain the mechanism by which the Updater writes contents variously to the belief box or pretense box. Similar considerations will apply, *mutatis mutandis*, to asymmetries in action production.

Finally, appealing to desire effects in order to keep within the single code/distinct attitude framework also has the rather different (and rather more pernicious) result of undermining the distinct attitude hypothesis. Nichols' proposal assumes that beliefs and imaginings are never, strictly speaking, treated differently by the affect mechanisms. The thrust of the desire effects proposal is that variant effects are due to variant desires. Applying this proposal generally would mean that, strictly speaking, no cognitive mechanisms treat beliefs and imaginings differentially. Perhaps this is an inevitable result of adopting the single code hypothesis. But *prima facie*, if this *is* the view, *we no longer wind up with functionally distinct attitudes in any substantive sense*. After all, it's

[12] I'll flag two possible objections here, without addressing them in detail. One: perhaps *subdoxastic* desires feature prominently, in which case the desires I deny are present may be there at an unconscious level. Some imagination-based accounts of dreaming, for instance, seem to suggest that unconscious desires (sometimes) direct the imagination (McGinn, 2004; Ichikawa, 2009). Two (relatedly): perhaps there's a 'competing desires' story to tell, wherein all cases of imaginational inference are appropriately related to 'victorious' (and perhaps subdoxastic) desires to imagine. Perhaps, but since in this context it's unclear whether there's any independent motivation for appealing to subdoxastic and/or competing desires other than simply to save the desire effects proposal, at present I won't consider these moves any further.

trivially true that the belief that *p* occurs in the same code as the belief that *p*. But *that* belief may very well produce different cognitive effects, given different sets of associated desires. The associations of beliefs to desires and imaginings to desires are highly contingent and variable. But this doesn't warrant the position that when different tokens of the belief that *p* have different effects (due to different sets of associated desires) those tokens are of different psychological kinds. What we need for the distinction in kinds seems to be differential treatment by a suite of cognitive mechanisms – e.g. compare the relation of belief to desire. But this – i.e. differential treatment by a suite of mechanisms – seems to be just what the new cognitivist has been forced to deny in order to keep within the single code framework. To be sure, there are subtle issues here regarding how to individuate cognitive kinds and cognitive processes. Nevertheless I think there's a real worry lurking here about whether appealing to desires in the proposed manner is a legitimate move for the new cognitivist.

In sum: I take the simple, labeled origins, and desire effects replies to fall short as reactions to the asymmetry challenge. No doubt the arguments provided in rejecting these replies fall short of being *decisive*. Nevertheless – absent some strong arguments to the contrary – direct appeals to the functional status of imaginings, appeals to thin imagination labels, and appeals to desire effects do not appear to promise us an adequate answer the asymmetry challenge for new cognitivism.

Before wrapping up this chapter, there are two rather more general objections to the challenge that need to be addressed. First, one may wonder whether the asymmetry challenge is fundamentally confused: Have I somehow misconstrued how cognitive

explanation is supposed to work? Second, one may worry that the asymmetry challenge *shows too much*: Are there asymmetry challenges for belief and desire, for example? For *every* propositional attitude?

6. Does the Asymmetry Challenge Confuse the Nature of Cognitive Explanation?

I anticipate the following reaction:

> You say that we require explanations for why action production mechanisms fail to consume imaginings, how the Monitor distinguishes imaginings from beliefs, how the Updater non-randomly writes to the belief box and pretense box, why the Inputter-Elaborator produces only imaginings, etc. But this is confused. The explanation for such asymmetries is *built into* the framework of cognitivist explanation. Cognitivists are functionalists, and categorize psychological kinds by direct appeal to their causal properties. There's nothing wrong with saying, for instance, that the reason that the action production mechanisms don't consume imaginings is just because *it's simply not part of the functional role of imaginings* to be consumed by action production. This is – as it were – all there is to it.

This amounts, I take it, to a retreat to the simple reply, and I think it simultaneously raises and misses the following important point: At a high enough level of explanatory abstraction – e.g. say, the level of functional explanation conducive to boxological depictions of the belief/desire/imagination system – the explanatory details of how the various mechanisms within the system work are lost. It's not that it's illegitimate to theorize at this level in general. In many cases, we can say all we need to say at this level. But describing the belief/imagination asymmetries is not one of these cases. If we're going to understand these cases, we need to – as it were – peek inside the boxes to see why and how the subserving mechanisms come to treat the relevant representations differently. One way to understand the asymmetry challenge is as the project of filling in these details in a way consistent with the new cognitive theory. I'm arguing the prospects don't look promising.

Actually, I think that this line of objection manifests an illicit slide between two modes of explanation – the one metaphysical, the other causal/cognitive. On one reading, the sense of the worry seems to be that the explanations for some asymmetries are just given by the facts about *what it is to be* a certain kind of psychological state. In developing the architecture, one might think, we can simply read off of the nature of imagination. E.g. It's subject to the will, so there must be mechanisms for that (the Inputter-Elaborator), it doesn't directly produce action (so it must be cut off from action production mechanisms), and so on. But facts about the deep nature of imagination are the wrong kinds of facts to appeal to in answering the asymmetry challenge. In many cases, these deeper facts are just descriptions of how imagination appears to differ from

belief. The asymmetries challenge requires us to develop a cognitive architecture which demonstrates how the functional properties characteristic of the relevant psychological kinds are realized. It may be, for example, that bearing a proprietary link to intention-direction is part of what it is to be an imaginational state. But it remains to be accounted for how that direct link is realized in us: what are the peculiarities of our cognitive architecture, such that we can imagine that *p* more or less at will? New cognitivism, to its credit, provides some details about these peculiarities. I've argued that, unfortunately, the details provided make the asymmetry challenge a genuine problem for the approach.

7. *Does the Asymmetry Challenge Show Too Much?*

The following considerations lead to another important question:

> If sound, your arguments surely show *too* much. For it would seem that exactly analogous considerations apply equally well to *all* functionally individuated psychological kinds. In which case, it would seem that functional asymmetries *in general* are not explainable within a single code framework. But – and here's the worry – the existence of a single cognitive code is not only an assumption of *new cognitivism about the imagination* – it's a widespread commitment of

cognitivism, *per se*. For, so much is minimally required in order to account for the causal interactions across all the representational kinds.

Take Goldman (2006, p. 46), for instance, who metaphorically states this widespread (if largely implicit) commitment:

> 'Consider the case of desire and belief. Desire representations and belief representations should also share the same code. Otherwise, how could desires and beliefs "talk" to one another, which they have to do when a person executes practical reasoning?'[13]

In short, if the above arguments support my claim that *new cognitivism* is explanatorily deficient, then exactly similar arguments support the broader position that *cognitivism, tout court,* is explanatorily deficient. In which case, my arguments must have gone wrong.

To see why this objection is ill-motivated we need to clarify what's at issue. What we essentially have is a (putative) dilemma, which is supposed to derive from a pair of

[13] I take Goldman out of context here. Goldman's concern regards whether the single code hypothesis is adequate to explain even commensurate psychological processing. Presumably, he suggests, beliefs and desires occur in the same code, but this doesn't imply that they'll be subject to commensurate processing.

(*prima facie*) inconsistent assumptions. The first assumption is the lesson culled from the asymmetry challenge; the second assumption is widely held:

(*A1*) Explaining functional *asymmetries* across tokens of distinct psychological kinds requires *denying* that they occur in a single cognitive code (i.e. denying that are semantically and structurally isomorphic).

(*A2*) Explaining the causal relations between tokens of distinct psychological kinds requires *assuming* that they occur in a single cognitive code.

Given the *prima facie* inconsistency between (*A1*) and (*A2*), it looks as though I am led to an unhappy dilemma:

(*C*) Either the asymmetry challenge is not genuine, or there's something wrong with a central feature of cognitivism generally.

Either way, the asymmetry challenge has misfired.[14]

[14] The dilemma formulated here might be adapted into a more aggressive objection to the asymmetry challenge roughly as follows: (*P1*) If the asymmetry challenge is genuine, then there is no single cognitive code for mental representations of distinct kinds; (*P2*) There must be a single cognitive code for (at least some) mental representations of distinct kinds; (*C*) So, the asymmetry challenge is artificial. I take it that

I maintain, however, that (*C*) is false – while the asymmetry challenge is genuine, this implies no general problem for cognitivism. The reason is that the force of the dilemma turns on an equivocation on the term 'single code'. One moral to take from the discussion above is that there are (at least) two senses of 'single code' to disambiguate. The term 'single code' should be read in a comparatively strong sense in the context of the single code hypothesis, where representations of distinct functional kinds (whose contents are picked out by the same *de dicto* clause) occur in a 'single code' in the sense that they are *compositionally* isomorphic.

But while this is one sense of the term, it is *not* the notion that is assumed by cognitivism more generally. When Goldman, for example, maintains that beliefs and desires occur in a single cognitive code, he's not committing himself to the view that the belief that *p* and the desire that *p* are compositionally equivalent. Rather, he is merely endorsing the weaker thesis that the two representations are constituents in a *unified representational system*. Accordingly, beliefs and desires occur in a 'single code' in the sense that they are permitted to interface with one another and with a shared suite of cognitive mechanisms. But this alone does not require assuming additionally that the belief that *p* and desire that *p* are compositionally isomorphic. After all, distinct representations of the *same* kind – say, the belief that *p* and the belief that *q* – presumably must also occur in a single code if they are to causally interact with each other. But no

---

my replies to the somewhat weaker dilemma version will apply equally to this stronger version of the objection.

one would suppose that, for this reason, these two beliefs must be compositionally isomorphic.

In fact, over the years, cognitivists have tended to avoid asymmetry challenges altogether by developing models which explicitly reflect a commitment only to the more inclusive sense of 'single cognitive code'. Take another – rather dusty – example from the history of classical planning systems: those employing the **STRIPS** language (the **ST**anford **R**esearch **I**nstitute **P**roblem **S**olver; Fikes and Nilsson, 1971). **STRIPS** incorporates representation-types including *initial states*, *goals*, and *commands*. What bears emphasizing for present purposes is that tokens of each type encode their own functional role – e.g. '*Init(At(Flat, Axle) ^ At(Spare, Trunk))*', '*Goal(At(Spare, Axle))* ', and '*Action(Remove(Spare, Trunk))* '. As a result, **STRIPS** representations of different kinds are not – at least arguably they are not – compositionally isomorphic. However – and this is the point – positing compositional differences across representation kinds provides no reason to assume that they fail to occur in a single cognitive code. There's nothing very special about **STRIPS** *qua* cognitivist model in this regard.

In sum, the asymmetry challenge does not show too much, and the suspicion that it does relies on an illicit equivocation on the term 'single code'. Moreover, there appears to be no truly worrisome version of this objection. Should one correct the equivocation by adopting the strong reading of 'single code' across (*A1*) and (*A2*), the dilemma would indeed appear to draw the legitimacy of the asymmetry challenge into question. However, as noted, on such a reading (*A2*) would be patently false, since cognitivists needn't

embrace the strong sense of 'single code' as a general commitment. But should one correct the equivocation by adopting the weaker reading of 'single code' throughout, the dilemma would wholly miss its target. For, the asymmetry challenge is only supposed to apply where robust functional asymmetries (at least of the gauge required by the distinct attitude hypothesis) are exhibited by representations which are supposed to be causally very similar *and* compositionally isomorphic. This is also why there's no asymmetry challenge for belief and desire. Even if we assume that beliefs and desires are compositionally isomorphic, the challenge is ameliorated by the fact that beliefs and desires are never treated similarly by any overlapping cognitive mechanisms. Beliefs and desires thus clearly causally distinct in a way which beliefs and imaginings, according to the new cognitive theory, are not.

8. Conclusion

In this chapter I've argued that the asymmetry challenge poses a serious explanatory problem for new cognitivism. The most plausible responses to the challenge are unsatisfactory. In view of this, we have good reason to suppose that new cognitivism paints itself into an explanatory corner. Positing a compositional isomorphism between beliefs and imaginings may help explain commensurate processing, but it leaves little in

the way of resources for explaining the systematic instances of incommensurate processing which the distinct attitude hypothesis demands.

Where does this leave cognitive theorizing about the imagination? If my diagnosis of new cognitivism's failure is correct – i.e. if the problem arises from an overly strict adherence to the single code hypothesis – then a plausible alternative response is to posit a subspecies of representation individuated by a characteristic *compositional* profile.

This style of account, as I've noted above, has precedents. Moreover, very different possible versions are currently being explored (Leslie, 1987; Langland-Hassan, 2011). These accounts suggest a promising way out of new cognitivism's explanatory corner, moving forward by rejecting the single code hypothesis, and thus undercutting the primary motivation for introducing propositional imagining as a *sui generis* psychological kind. For his part, Alan Leslie (2002, p. 105) provides an illuminating restatement of the general thrust of this alternative sort of approach, which we'll move to explore further in the remaining chapters:

> 'Although in accounts of behavior there are usually trade-offs between process and representation that make their effects hard to distinguish, sometimes it is possible to distinguish between the two…[T]he case of *belief* versus *pretense* provides such an example. We find that the ability to *have a belief* rests on a mode of processing…By contrast, in the sense corresponding *to having a belief*, there is no such thing as a *having a pretense*, nothing for the case of pretense that

rests on a special mode of processing. Instead, the ability to pretend rests on a special representation.'(original emphases)

The asymmetry challenge and the considerations here lends some motivation and *prima facie* support to these most general assumptions behind this evolving alternative to the default approach to understanding the propositional imagination.

Chapter 2: The Single Attitude Account of Imagination

'…[T]he activity of imagining that p *consists merely in retrieving one's beliefs in generalizations relevant to the proposition that* p, *and using them to make judgments about what would likely happen if* p...' (*P. Langland-Hassan, 'Pretense, Imagination, and Belief: The Single Attitude Account', 2011*)

In this chapter I present and evaluate the *Single Attitude Hypothesis* (*SAH*) (Langland-Hassan 2011). *SAH* presents a novel account of pretense and imagination, according to which the primary vehicles of imagining that *p* are beliefs with counterfactual conditional contents. *SAH* thus marks a sharp and novel departure from the distinct attitude approach taken by proponents of the new cognitive theory.

My purpose here is to argue that, while *SAH* provides a viable account of the mechanisms and role of imagination in pretense, the proposal is incapable of telling the whole (psychological) story of imagination. For, once we begin to consider the role of imagination in accessing varieties of fictional contents, it becomes evident that the *SAH* architecture is too austere to capture the range of the propositional imagination.

1. The Single Attitude Hypothesis

The central strategy behind *SAH* is to explain pretense recognition in young children (more of which in chapter 4) and pretense behavior by exploiting the intuitive relationship between the propositional imagination and counterfactual reasoning. The thrust of the proposal is that the imaginational representations and mechanisms underpinning pretense are just those which underpin counterfactual inference generally. This assumption is then marshaled in favor of a pair of interesting positions.

First, it's argued that pretending that *p* doesn't require knowledge or concepts of the psychological states of pretenders. *SAH*, thus purports to *settle* a long-standing debate in the cognitive science of pretense. I maintain that *SAH* fails on this point, but here I wish to focus specifically on the account of imagination that accompanies the account of pretense. Hence, I'll save my arguments on this point for later (chapter 4).

Second, it's argued that pretending that *p* doesn't require any special, *sui generis* attitude of imagination. To be clear, however, this is not an *elimination* argument with regard to the imagination, as if eliminating the *sui generis* attitude would be to eliminate the capacity or the psychological kind we refer to by 'imagination'. Rather, the general motivation behind *SAH* is that the distinct attitudes of the new cognitive theory are explanatorily expendable. In keeping with a long-standing tradition in imagination theory – Langland-Hassan opts to take the role of imagination in early pretense as a critical case for displaying how we can get by without the kinds of imaginings described by the new

cognitive theorists, on the one hand, and without those described by the metacognitivists, on the other.

Here is the core idea: To imagine that *p* – e.g. to imagine *that it's snowing on Denali* – effectively involves the imaginer in asking herself, 'What if it were snowing on Denali?', before subsequently reasoning about what *would be* the case, *were* it snowing on Denali. This involves an imaginer in representing contents such as *that if it were snowing on Denali, then snow would be accumulating on the South Summit*, and so on, as derived from the imaginer's more general knowledge of Alaska, snowfall, mountains, Denali, etc.

But what this is never supposed to require is representing the content *that it's snowing on Denali*, full stop, and in some functionally distinct component of the architecture, e.g. in a pretense box, in a possible worlds box, in quotation marks, or whatever. We've seen in prior chapters how the latter sorts of account look in practice. Let's see how the *SAH* alternative is supposed to work.

SAH earns its keep by furnishing a comparatively elegant account of pretense recognition and behavior. There are essentially three stages to the process. First, pretend play is identified when the subject acknowledges that a special kind of activity – i.e. a kind of *pretense game* – is engaged by a playmate (this stage might be bypassed when the subject is the initiator of the game). Then, if appropriately motivated, the subject engages in the game herself by representing the perceived counterfactual conditions of the

pretense. Finally, the subject proceeds to behave *as would be appropriate if* those conditions obtain.[15]

Here's a well-worn example: A toddler drawn into a game of tea parties learns by observing Mother that a special kind of activity has been initiated – one in which Mother behaves *as would be appropriate if we were at a tea party*. The motivation to join in the game causes a search for further relevant information to initiate. The information search, in turn, initiates a cascade of cognition drawing upon relevant knowledge.

This process would look something like this:

(Belief) MOTHER STARTS A GAME IN WHICH WE ACT AS IF WE ARE AT A TEA PARTY

(Desire) I PLAY THE GAME

(Search) WHAT IF WE WERE AT A TEA PARTY?

So long as there is sufficient, relevant knowledge at the ready, the young pretender may then track and contribute to the direction of the pretense by attending to salient behavioral cues:

[15] It's worth observing that distinct attitude theorists endorse a similar account of pretense recognition, eschewing the idea that pretense recognition requires that young pretenders to understand that their playmates imagine or pretend that *p*. *SAH* is thus presented as kind of an intermediate option, capturing the data without distinct attitudes, but also without illicitly imputing sophisticated psychological concepts to toddlers – more of which in chapter 3.

(Belief) MOTHER ACTS AS IF POURING TEA

These kinds of cues (as always, together with the right motivational states and background knowledge) are supposed to be sufficient for stimulating the further imaginative activity required to successfully extend engagement with the pretense:

(Belief) (*x*) IF *x* POURS LIQUID$_{L}$ INTO CUP$_{C}$, THEN CUP$_{C}$ IS FULL

(Belief) IF MOMMA WERE TO POUR TEA$_{L}$ INTO CUP$_{MY}$, THEN CUP$_{MY}$ WOULD BE FULL OF TEA$_{L}$

The cumulative effect of the token beliefs and desires is that the subject, too, behaves as if she were at a tea party. Since she knows what to do, for instance, with cups full of tea, she will go ahead and act as if taking a sip, and so on.

The two crucial pieces of this sketch of the process envisaged by proponents of *SAH* are that (*i*) the process does not involve referring to the mental states of Mother – e.g. as in that Mother pretends or imagines that *p*, and that (*ii*) the process doesn't involve

representing any propositions in the functionally distinct (cognitive) workspaces described by both metacognitivists and new cogntivists.

Thus, the virtues of *SAH* come in its promise of economy and parsimony. *Contra* metarepresentational accounts, *SAH* doesn't require us to attribute any psychological concepts to young pretender/imaginers. But *SAH* also eschews much of the imagination-specific machinery the new cognitive theory requires us to adopt. The chief architects (Nichols and Stich, 2000, 2003) of new cognitivism are led to propose, for example, that the interface between knowledge and imagination requires that much or all of the contents of the belief box are *copied into* the pretense box (with the elimination of explicit contradictions) at the start of a pretense game. Moreover, as a pretense is elaborated in the imagination box, the authors propose a *synchronous unfolding* of counterfactual contents in the belief box – contents such as *if Mother were to pour the tea, then my cup would be full and I might drink from it*. And it's *these* counterfactual beliefs that are supposed to function in the proximal causation of pretense appropriate action. There's little said about why there should be such a synchronicity, or how it is achieved.

SAH thus promises to eliminate much of the auxiliary architecture the distinct attitude view requires us to take on board. There's no need to propose that the contents of the belief box are copied into a distinct imagination box – by hypothesis, imagining *occurs* in the belief box, and thus (in principle) any belief stored there is capable of interfacing with imaginational contents. Moreover, there's no need to explain the relation

between representing *p* in an imagination box and representing (the motivationally relevant) *if p were, then q would be* in the belief box. For, the imagined content is embedded in the motivational state, and is *a fortiori* tokened in the belief box.

No doubt there are flags which ought to be raised here. For instance, as it's presented, *SAH* has nothing much to say about *how* inferences from background knowledge to the specifically *counterfactual* beliefs identified with imagining that *p* are executed. The capacity to move from ordinary beliefs to counterfactual, or subjunctive, beliefs has simply been taken for granted.

Moreover, one may find the prediction that imaginational inference in pretense proceeds without representing a content *p* – full stop – counterintuitive. How do we get to infer counterfactual conditionals, after all, if not by *independently* representing their antecedents in imagination? How do we produce independent representations of the consequences? If there are cases where the antecedent *p* is represented prior to the production of a corresponding counterfactual representation, it looks like this would be enough to save the received view. For, this would *prima facie* involve representing *p* in the imagination role (or box), before linking it to its putative consequences in the role of counterfactual belief (Harris, 2000; Currie & Ravenscroft, 2002; Nichols & Stich, 2000, 2003).

I propose to hold off on addressing these kinds of issues until I have a chance to sketch out my own, pluralistic proposal in chapter 4. My aim here is rather to show that

there are some – that there are *lots* – of imaginational phenomena which *SAH* cannot accommodate.

One caveat before moving to the critical portion*:* As noted, *SAH* is developed explicitly as an account of pretense recognition and causation in conceptually naïve toddlers (i.e. naïve with respect to other minds). Nevertheless, Langland-Hassan (2011, p. 3) suggests that *SAH* might be extrapolated into a fuller account of imagination. For instance:

> 'My view, in a nutshell, is that imagining that *p* amounts to making judgments about what would likely happen if *p*, from retrieved beliefs in generalizations…'

For present purposes I will simply treat *SAH* as potentially wide in scope – i.e. as a general theory of propositional imagination, to include an account of pretense. In doing so, we'll see, there are cases which cast considerable doubt on whether *SAH* could provide a comprehensive framework for understanding the propositional imagination.

2. *Problem Cases for SAH*

While the cases I'll submit here multiply, and vary greatly in their details and potential manifestations, they're unified in challenging the central tenet of *SAH*: That imagining that *p* ordinarily involves us in producing representations with counterfactual conditional

contents. I'll argue rather that, *prima facie*, a good deal of imagining seems to eschew counterfactual reasoning altogether.

2.1 Semantically Rich & Semantically Poor Imagining

Representing fictional worlds should involve us in imagining, if anything does. But many fictions are so rich in the contents they deliver to imagination that detouring into counterfactual reasoning in order to track their contents scarcely seems necessary.

To be clear: It's not that we do not ordinarily make counterfactual inferences about the rich fictional worlds we engage – more of which below. It's just that, as an apparent rule of thumb, the richer the fictional world described, the less we tend to actively contribute to that world over the course of our engagement.

Consider, for example, the world delivered to us in the novel *Jane Eyre*. In this case, Bronte (1847) weaves a complex world for the consumer's imagination. Here's a taste:

> 'There was no possibility of taking a walk that day. We had been wandering, indeed, in the leafless shrubbery an hour in the morning; but since dinner (Mrs. Reed, when there was no company, dined early) the cold winter wind had brought with it clouds so sombre, and a rain so penetrating, that further out-door exercise was now out of the question. I was glad of it: I never liked long walks, especially

> on chilly afternoons: dreadful to me was the coming home in the raw twilight, with nipped fingers and toes, and a heart saddened by the chidings of Bessie, the nurse, and humbled by the consciousness of my physical inferiority to Eliza, John, and Georgiana Reed. The said Eliza, John, and Georgiana were now clustered round their mama in the drawing-room: she lay reclined on a sofa by the fireside, and with her darlings about her (for the time neither quarrelling nor crying) looked perfectly happy…'

The question before us is: We can assume that imagination plays a critical role in consuming fictions of this kind – but what role, if any, does counterfactual reasoning play?

According to the mechanisms of imagination described in *SAH*, imaginatively engaging this fiction should involve us in initiating deliberative searches over questions like, e.g. 'What if there were no possibility of taking a walk that day?', or, 'What if Mrs. Reed were to dine early?'. *Prima facie*, this is altogether unnatural as an account of what is going on under the surface as the world of *Jane Eyre* begins to take shape in our imaginations. Moreover, when we do engage the contents presented by asking such questions, (again, *prima facie*) we are doing something more than simply tracking the contents of the fictional world delivered. Indeed, (one might've thought) such richly described fictions rather *invite the suspension* of deliberate reasoning.

Imagining semantically impoverished fictional worlds seems likewise only contingently related to counterfactual reasoning. Here, for instance, is a (very) short story, which I've just come up with. It's called *Chicken Farmer*:

The chicken farmer ate breakfast.

Chicken Farmer certainly is an impoverished piece of fiction. Nevertheless, we presumably involve the imagination when we represent its contents. Moreover, *prima facie*, we (can) do so without any motivation to address questions like, 'What if the chicken farmer were to eat breakfast?'

When we do engage the story, *however* we do it, it's not at all natural to assume that what we're doing involves the sort of counterfactual inferences *SAH* assimilates to imagining that *p*. To a first approximation, rather, representing such sparse contents seems to involve us in producing what we might call 'bare imaginings', the production of which is supposed to map roughly onto the psychological category we might associate with merely entertaining that *p* (more of which in chapter 4).

This is not to say that we're not representing a lot more than the single proposition – barely imagined, as it were – which describes the world of *Chicken Farmer*. For example, there's likely to be associated mental imagery, which may lead to the representation of further imagery and perhaps further propositions associated with the

images conjured. But none of this seems to require that we reason in the specifically counterfactual, or subjunctive, mood in order to imagine *that the chicken farmer ate breakfast*. It seems that we have the cognitive wherewithal to entertain such propositions in the mood of bare imagination.

Of course, even the richest of fictions doesn't (explicitly) specify all of the properties of its world – probably not even all of the interesting or important properties. This fact about fictions ensures that there will be plenty of work for the counterfactual reasoning mechanisms to do when it comes to imagining fictions. Part of what makes *Jane Eyre* so engaging are the psychological contrasts surrounding its characters – e.g. Jane, Sarah, and Eliza. A lot of these contrasts are derivable only from the subtext (by definition inexplicit) created – either intentionally or unintentionally – by Bronte. Accessing this subtext may (at times) require counterfactual reasoning over the characters' utterances and other behavior. Nevertheless, *prima facie*, such detours constitute an auxiliary activity. They involve a deliberative process (though perhaps not one that needs to be consciously or voluntarily initiated) of counterfactual reasoning which brings additional, but inessential, contributions from imagination.

Similar considerations will apply, *mutatis mutandis*, across fictions and genres in which artists succeed in delivering compelling characters and plotlines.

2.2 Bizarre Fictions & Flights of Fancy

Bizarre imaginings and random flights of fancy constitute another problematic family of cases for *SAH*.

Generally speaking, it's easy for us to imagine contents which range from just a little weird to as bizarre as you like.[16] Here's an excerpt from another story of mine, called *Nebraska Jack*. It starts out like this:

> It was Tuesday morning in Axtell, Nebraska, and Jack was up especially early. Ordinarily, he was accustomed to rising *with* the Sun rather than before it, but today the dogs awoke him just prior to sunrise. The small deviation from his routine didn't bother Jack, and he moved on to his usual breakfast of six dozen donuts and an oil drum of coffee…

We can gather that there's something very odd about Jack, though lots of what occurs suggests that Jack's world is very near to our own in many respects. The odd part is, of course, that Jack consumes such an extraordinary breakfast.

[16] There are important exceptions to this generalization reviewed in the extensive literature on imaginative resistance. It may be hard to imagine bizarre propositions like, for instance, *that Macbeth was morally praiseworthy for murdering Duncan* (Moran 1994), or *that 5 + 7 ≠ 12* (Nichols 2004). I assume that these kind of exceptions don't affect the general truth of the claim that imagining bizarre propositions is easy. We seem to have no trouble, for instance, imagining *that in a different galaxy, and in the distant past, there were intelligent robots with British accents*.

The question for us is how we should *so easily* imagine the odd part, *if* doing so is a matter of drawing out consequences based on our background assumptions (in accordance with *SAH*). Actually, backing up a step, there's a pair of relevant questions here: First: Does imagining the contents of Nebraska Jack involve deliberating on questions like, e.g. 'What if Jack were to eat six dozen donuts every morning?' Second: How do the counterfactual reasoning mechanisms produce imaginational representations which deviate so sharply from our background assumptions about reality? No one eats that much for breakfast, but (by all accounts) the mechanisms of counterfactual belief production aim at coherence with background beliefs.

I addressed the first type of question above, and similar considerations apply. As in the cases of *Jane Eyre* and *Chicken Farmer*, it's *prima facie* implausible to suppose that we *must* involve ourselves in deliberating over such questions in the course of imagining the contents of *Nebraska Jack*. There may be reasons to assume that the process is mostly unconscious, but barring some convincing arguments from the *SAH* camp, this seems like a stretch.

The second question – How are reality-deviant counterfactuals produced? – presents a different sort of challenge. I suppose one might say that *just the fact of noticing* that Jack is odd points to the involvement of counterfactual reasoning mechanisms. It seems to involve, at least, noticing that some contrary-to-ordinary properties obtain. Maybe this noticing itself involves drawing inferences to contents such as that *if Jack were to eat six dozen donuts, then he'd be very strange*.

Perhaps. But these sorts of judgments apparently come *after* imagining the contents of the story. And, moreover, a psychologically normal consumer of fiction might nevertheless engage *Nebraska Jack* without ever having surmised that Jack is strange. In which case, it remains unclear whether and what sort of counterfactual reasoning process would be involved in – much less constitutive of – imagining (for instance) *that Jack eats six dozen donuts for breakfast every day*. According to *SAH*, imaginings are essentially derivations from our general knowledge about the world. But the prescribed imagining in *Nebraska Jack* seems rather to require the *suspension*, or perhaps the temporary modification, of background assumptions.

Moreover, as a general rule: The more bizarre the fiction, the more tenuous the relation between imagination and background knowledge. I recall a daydream I once had I'll call *Alien Invasion*. As I remember, it unfolded something like this:

> I'm riding my bike home from the library. As usual, I take a left onto the Lane Avenue Bridge when – lo and behold(!) – a legion of well-armored, alien soldiers pours out of the Schottenstein Center toward me. I freeze. As they surround me, one of them disappears, before promptly re-appearing inches in front of me. The creature steps forward and offers me a Snickers bar, along with a certificate for a year's worth of dry cleaning.

Flights of imaginative fancy like this one strike me as fairly routine exercises of imagination. But, *prima facie*, they're very far from what counts as any routine exercise of counterfactual reasoning. In particular, it's hard to see how such strange contents could be plausibly *retrieved from* or *derived from* stored beliefs, in accordance with the account delivered by *SAH*.

By contrast, flights of fancy apparently eschew deliberative epistemic processes altogether – this is a regular part of the content production and phenomenology of involuntary daydreaming. Let's suppose, for the moment, that the *Alien Invasion* daydream kicked off with a search for information *a la SAH*: e.g. 'What if I were riding my bike home from the library?' There doesn't seem to be any plausible reason to predict that the contents elaborated in imagination would come out anything like *Alien Invasion*. The story would've been rather more boring, detailing my normally uneventful passage from the library to my apartment.

2.3 Taking Stock

In the case of rich and poor fictions, it's unclear whether and to what extent we engage in counterfactual reasoning in order to imagine the contents prescribed. And if we do not, then we need an auxiliary account of the architecture of propositional imagining in such cases. Thus, what we've seen so far is that *SAH* does very well in handling cases of imagination and pretense where the contents conform closely to the subject's background

knowledge. However, accounting for the production of bizarre imaginational contents presents a *prima facie* problem.

On reflection, if it should turn out that the relation between counterfactual reasoning and imagination is contingent, rather than constitutive, this would be sort of unsurprising. For, the imagination has long been assumed to evade (or at least sometimes to evade) constraints of normativity and relevance which govern belief – counterfactual or otherwise. Presumably, this is part of the reason *why* it's so comparatively easy for us to imagine that Nebraska Jack consumes such an unusual breakfast, and that an alien soldier wants me to have candy and free dry cleaning.

Nevertheless, *SAH* presents an intriguing alternative to the usual strands of metacognitivism and new cognitivism about pretense and imagination. Thus, it's worth exploring some potential objections and replies on behalf of the *SAH* proponent. Ultimately, I don't think the challenges I raise here are surmountable within the *SAH* framework. Furthermore, I raise further challenges for the accompanying account of pretense recognition in chapter 3. However, I maintain, none of this is to say that the single attitude alternative should be *abandoned*, as it may be that it can be incorporated into a broader account of imagination.

3. *Objections & Replies*

Objection 1: It's unfair to criticize *SAH* for not providing a full account of imagination, including an account of how we represent the varieties of fictional contents described. After all, *SAH* is specifically designed to handle cases of imagination in pretense, and especially those in which children model various aspects of reality in playing pretense games.

Reply 1: Langland-Hassan at least gestures at the claim that *SAH* has the resources to give a fuller account of imagination. My aim has been to delimit the scope of *SAH*. In the process I've shown where some of the work of challenging available accounts remains to be taken up – e.g. in imagining rich, poor, and bizarre fictions.

However, we must also observe that what goes for imagining rich, poor, and bizarre fictions apparently *also* goes for pretense. While *SAH* focuses on simple pretense games involving the production of counterfactual, cognitive models of more or less realistic scenarios, it's clear that pretend play is not so limited. Episodes of pretend play can be as rich, poor, or bizarre as you like. In such cases, we're going to need extensions to the *SAH* proposal which capture imagining beyond the scope of counterfactual belief; for, objections which exactly parallel those I raise with respect to fictional contents above will apply the analogous varieties of pretense contents.

Objection 2: Rich and poor fictions to one side, why *can't SAH* accommodate bizarre imaginings and flights of fancy? Take another example directly out of the original articulation of *SAH* (Langland-Hassan 2011, my emphasis):

> 'If you wish to pretend that a tornado strikes at [a] tea party, you use stored generalizations about tornados to reason about how they would affect a tea party. The freedom of imagination is just a special case of the more general freedom we have to reason about topics of our own choosing. *Sometimes this involves reasoning about the likely consequences of scenarios that are themselves deemed unlikely or unusual*…'

A tea party is admittedly a rather mundane piece of reality. But a tornado-struck tea party is far from ordinary. Exactly where is the problem for *SAH*?

Reply 2: The central idea here is to treat bizarre turns of imagination and pretense as deliberate, but bizarre *interventions* in chains of counterfactual reasoning. In the example just rehearsed, for instance, the subject interjects contents into the typical tea party script by asking, 'What if a tornado were to strike my tea party?' She then has only to draw on her background knowledge in order to fill in the relevant scenario.

There are a several points to make about this suggestion. First of all, there is definitely something attractive about the suggestion that the doxastic voluntarism that seems to accompany imagination (i.e. it's sensitivity to intention direction, as I put it in

chapter 1) is nothing, really, over and distinct from our general capacity to deliberate over the topics we choose. Why should it be so mysterious – or rather, *more* mysterious – that we can imagine that *p* more or less at will, when we can deliberate over *p*, *q*, *r*, or whatever else when we choose. I think this insight surely points to a measure of progress in our understanding of the challenges of accounting for the psychological mechanisms of imagination. In this case, it does not look as though we need to understand the relevant phenomenon as a distinctive causal property of a distinctive psychological kind at all.

For a second point: this probably *is* the way bizarre imaginings get produced sometimes. I think I can admit, that is, that interventions like tornado-at-a-tea-party cases are an entirely natural way one might spice up a pretense game. Or, for that matter, it seems like a legitimate means to brainstorming counterfactual scenarios generally.

Unfortunately, this is not really an answer to the challenges I raise above. For, it remains unclear whether we should be required to *deliberately* interject interventions of this sort over the course of our elaboration of fictions in imagination – be they bizarre or ordinary. The response we're presently entertaining seems to be that, really, none of the contents needs to flow naturally from preceding contents. All we need to do is to continue to interject queries over contents that can be more or less bizarre.

But if this is the answer, then very bizarre cases, in particular, put a lot of pressure on it. Take *Alien Invasion*, for example, which is apparently constituted by a series of increasingly bizarre contents, none of which seem to flow naturally from others, and none of which seem to flow naturally from background knowledge. The response we're

entertaining seems to lead to the position that the whole scenario is really just one long string of interjected queries over bizarre possibilities. As if, for example, my daydream involved me in deliberating over, 'What if I were riding home…and I were accosted by alien soldiers…and they offered me candy…?', and so on. One might have thought, in fact, that the whole conjunctive query state would be but a prelude to the real imaginational process. At least that's how it looks in the tornado-at-a-tea-party case. At any rate, it should surely count as a strike against *SAH*, if its best option is to reduce such very strange detours of imagination to bizarre queries over possibilities.

Objection 3: The general thrust of the critique of *SAH* is in the claim that imagination and counterfactual reasoning cannot be fully assimilated (and thus, presumably, that the mechanisms of each must come apart). But this doesn't do justice to the complexities of the dependencies across the two. Imagining fictions – rich, poor, bizarre, and intermediate – surely always involves filling in inexplicit contents. We don't imagine, for instance, that the chicken farmer ate breakfast, *tout court*. We imagine that the farmer is a man (or a woman), that he/she is wearing clothing (probably overalls), and so on. This 'filling in' process is best explained as the work of the counterfactual reasoning mechanisms, and thus the connection between counterfactual reasoning and imagination must be tighter than the arguments presented here seem to suggest.

Reply 3: As I think I've at least suggested along the way herein, I'm happy to posit a very tight connection between the mechanisms of imagination and counterfactual reasoning. I have other reasons, which I'll make clear later (chapter 4) for thinking so.

Nonetheless, we need to be careful about how we understand the connection. Let's say, for the sake of argument, that mentally representing fictional contents *always* involves the imaginer in contributing to the contents represented. Now let's ask: What might this suggest about the role of the counterfactual reasoning?

It would be a mistake to assume that all the inferential (as it were) filling in that goes on while consuming (or, for that matter producing) a piece of fiction is accomplished *via* counterfactual reasoning. In the *Chicken Farmer* case, it is right to point out that we do not merely imagine that a farmer ate breakfast, but also, for instance, that we (probably) imagine the farmer as either male or female. However, the auxiliary representations produced are as likely to result from *pragmatic* assumptions and inferences as they are from counterfactual reasoning.

By 'pragmatic assumptions and inferences' here I mean something very general. Minimally, we should expect that the terms of the story – e.g. 'chicken', 'farmer', and 'breakfast' – and the concepts evoked – e.g. CHICKEN, FARMER and BREAKFAST – will activate a web of semantic relations and auxiliary assumptions by raw association. What, for example, does the term 'breakfast' bring to mind for you?

Moreover, in addition to associative connections, we should expect mental imagery to be accessed by default, to include (perhaps) images of a prototypical farmer – e.g. a middle-aged man in overalls – and a prototypical breakfast – e.g. two eggs with a side of meat. The point is not that any specific assumptions and/or images are accessed, it's just that some assumptions and images or others are more than likely to

spontaneously accompany the explicit content prescribed by the terms of the story. All this contributes to the fleshing out of what is not explicit in the story, and comes *in addition* to whatever 'filling in' role is played by the counterfactual reasoning mechanisms.

Objection 4: But hold on. If there's real substance to *Reply 3*, then what distinguishes imagining that *p* from the work-a-day (so-called) 'pragmatic assumptions and inferences' we make during language comprehension generally? Surely there's a role for what I've called 'pragmatics' in imagining that *p*. But on one reading it sounds like more or less any pragmatic assumption or inference could be counted as an imagining.

While reading *Chicken Farmer*, for instance, assumptions such as that the chicken farmer is a middle-aged man should surely be counted among our imaginings. However, suppose we're reading a passage about an actual chicken farmer in the *New York Times*. The same pragmatic mechanisms will produce similar representations and inferences, perhaps leading us to assume that the individual we're reading about is a middle-aged man. In this latter case, should the assumption be counted among our imaginings? Surely it would be more apt to say – given the context – that we believe that the chicken farmer referred to in the *New York Times* is a middle-aged man. While pragmatic mechanisms of content production of the general sort I mention are at work *all the time* when we're processing language, imagination arguably is not. By contrast, the production of content by the counterfactual reasoning mechanisms is much more plausibly assimilated to

imagination, *a la SAH*. Don't these considerations constitute *prima facie* reason to restrict the 'filling in' role to the counterfactual reasoning?

Reply 4: The first thing to say, perhaps, is that it's not clear to me that the imagination does not play a role in the consumption of non-fictions. It's not clear, for instance, that it should be worrisome to assume that an article in the *NYT* should invoke imaginings in us.

Setting this possibility to one side, I think we need to consider the important effects of mental imagery in relation to propositional imagining. On the one hand, we mustn't *confuse* mental images with propositional imaginings. Reading *either Chicken Farmer or* a *NY Times* article about a chicken farmer may equally evoke mental images of stereotypical chicken farmers. But mental images are, I take it, distinct from both ordinary beliefs and propositional imaginings. Even so, images may play an important role, *qua* images, in *adding* to the mental contents associated with our representation of both fictional and non-fictional states of affairs. In such cases, there shouldn't be any worry over whether the contents we report upon – e.g. that the chicken farmer is a middle-aged man – are beliefs or imaginings. For, we're simply reporting on mental imagery we're experiencing. That imagery may augment the total set of mental contents, to include propositional imaginings, ordinary beliefs, or both.

On the other hand, mental imagery may frequently play an important *causal* role in the inferential filling in of a fiction's inexplicit contents by facilitating the production of further propositional imaginings. When *Chicken Farmer* evokes images of a middle-

aged man in overalls, it may in turn produce beliefs with contents such as *that the chicken farmer in* Chicken Farmer *is a middle-aged man*. This belief is distinct from the mental image. It is also distinct from any specifically counterfactual (conditional) belief. Nevertheless, it should arguably be counted among the total set of imaginings making up our engagement with the story. Mental imagery may play a similar causal role when we engage non-fictions. None of this seems like it should lead to any deep worries over the functional status of the assumptions and inferences produced by pragmatic mechanisms of language comprehension.

4. A Dilemma and a Moral

No doubt a full assessment of *SAH* would require further discussion. What I hope to have shown is that the attempt to fully assimilate – i.e. to *reduce* – the architecture of the imagination to the architecture of counterfactual conditional belief production would probably be a non-starter. For, to insist otherwise would be to incur the burden of answering a rather cumbersome dilemma. The proponent of the proposed reduction would have to show *either* (*i*) that imagining rich, poor, and bizarre fictions *always* proceeds *via* counterfactual inference – i.e. by drawing upon background knowledge in order to link propositions together conditionally – *or* (*ii*) that rich, poor, and bizarre fictional worlds *not* represented *via* counterfactual inference fail to count as involving the imagination. Neither horn of this dilemma seems very plausible.

But the moral of the story here is certainly not that we should abandon *SAH*, or that we should otherwise give up on exploring a single attitude alternative to the new cognitivist framework. The moral is that we need a richer picture of the mechanisms involved in the production and maintenance of imaginings – one which captures, for instance, bare imagining, and perhaps other imagining which occurs outside the scope of counterfactual elaboration. Fortunately, there's ample room to develop this richer picture well within a parsimonious architectural framework.

Before I begin to develop this picture, however, there's a further point of contention in the literature to adjudicate – the question of whether metacognitive states are required for the recognition of pretense by very young pretenders. Interestingly, new cognitivists and single attitude theorists agree on an answer to this question. In broad strokes, they agree that – *pace* the metacognitivist approach – pretense recognition does not involve young pretenders in reasoning about the psychological states of their playmates – to include reasoning about the fact that they bear some proprietary attitude toward the contents of the pretense. Instead, a behavioralistic account of pretense is proposed, which requires that young pretenders attend only to available behavioral cues in the detection of pretense. As we'll see, *SAH* (and new cognitivism, for that matter) ultimately founder on this critical *desideratum*, as well.

Chapter 3: Why Pretense (Recognition) is Metarepresentational

'Does pretend play, when it emerges at around two years of age really require the child to represent another's counterfactual representations?' (C. Jarrold et al, 'Pretend Play: Is it Metarepresentational?' 1994)

A *prima facie* plausible assumption about competent pretenders is that they somehow track the fact that the contents of the pretend games they play depart from reality in various ways. Pretending that *p*, in effect, seems to involve imagining that *p*, or at least representing *p* in a way which does justice to the fact that *p* is *merely* the content of a pretense.[17]

This leads quickly to a pair of explanatory challenges. First, the obvious question: How do young pretenders pretend – what sort of representations subserve their pretense behavior? Second: How do young pretenders *detect pretense in others* – to include knowing that their playmates take the contents of a particular pretense to be unreal?

[17] In special cases it seems to involve representing that that *p* is *also* pretended – e.g. when *p* also happens to constitute an accepted part of reality.

For present purposes I wish to focus on the second question. For, the arguments over whether pretense recognition is metarepresentational – whether it involves tracking the imaginational and motivational states of other agents – or whether it is behavioralistic – whether it involves tracking only the observable behavioral cues of other agents – have recently resurfaced in the literature (Nichols and Stich, 2003; Friedman and Leslie, 2007; Friedman *et al*, 2010; Langland-Hassan, 2011).

Moreover, I maintain that the debate has reached a decisive point, with the appearance of the most sophisticated behavioralist account to date (Langland-Hassan, 2011). [18] My goal here is to argue that, in spite of a strong push to eschew metacognition in accounts of pretense recognition, we have yet to see a satisfactory behavioralist alternative. Minimally, it is difficult to see how the behavioralist account might work without at least a basic capacity for metacognition. Moreover, it's difficult to see how it could work without recourse to a notion of pretense that's very like the notion Leslie initially argued is required. Hence, *pace* the push for behavioralism about pretense, available considerations suggest that pretense recognition requires metacognition.

[18] It's perhaps worth emphasizing at the outset that no one involved here is a behaviorist, *per se*. All parties are cognitivists, in the business of developing representationally rich accounts of pretense and its subserving mechanisms. The disagreement here is a disagreement among cognitivists over whether pretense and pretense recognition involve metacognitive states. I reserve the term 'behavioralist' for accounts which eschew metacognition.

1. The Neo-Behavioralist Approach to Pretense Recognition

According to the neo-behavioralist approach (Harris, 1991; Nichols & Stich, 2003; Langland-Hassan, 2012), the mechanism of pretense detection works without the capacity to track the unobservable psychological states of pretenders. Instead, the mechanism tracks observable properties of *situations*. In effect, the idea is that sensitivity to pretense requires a sensitivity to *a special kind of game*, one in which the players act *as would be appropriate if p*, though *p* may not obtain.

To be clear, behavioralists don't deny that (even very young) children possess some concept of pretense. Rather, they just claim that the referent of that concept is a game of the kind just canvassed, and one that can be picked out without understanding that a rich psychological subtext is at work below the surface. Taking up with the strong behavioralist current in the literature, Peter Langland-Hassan (2011) has recently tried to settle the discussion of pretense recognition in favor of the behavioralist answer to the question of pretense detection.[19]

Langland-Hassan extends the behavioralist notion that young pretenders need only grasp that a particular (pretense) situation is one in which a playmate behaves *as would be appropriate, provided that p*. In the relevant contexts, the pretense detector is involved in attending to, and storing representations of, whatever conspicuous manner cues indicate that a pretense is in effect. Simultaneously, (conspicuous) cues that a

[19] His account is intended to compliment his single attitude account of imagination (rehearsed in chapter 2).

playmate behaves *as would be appropriate, were not-p* are also carefully tracked. Thus, the detector is supposed to be sensitive to situations in which the appropriate balance of *as if p* and *as if not-p* behavior obtain.

The crucial piece of Langland-Hassan's account lies in its novel attempt to articulate a thoroughly behavioralist notion of a *pretense game*. Recognizing pretense games, the story goes, is achieved by the detector's tracking of when a set of manner cues (as it were) add up in the right sort of way, thereby activating the (behavioralistic) concept GAME. The particular contents of pretense games are picked out simultaneously, as the cues inciting detection also indicate that some object(s) *x* is made to be like some other object (or kind of object) *y*. I'll rehearse the details of how this is supposed to work below.

This new account is presented as a direct reply to the metacognitivist charge that a behavioralistic pretense detector simply would not function properly. Specifically, Leslie and Friedman (2007) and Friedman *et al* (2010) argue that behavioralist accounts are both too *inclusive* of the wrong sorts of behavior, and too exclusive of the right sorts of behavior. In short, the charge is that the behavioralists describe a mechanism which ultimately fails to track the relevant property in the world – the property of *being a pretense game*.

To see why, first notice that behavioral accounts *prima facie* predict that children should categorize a lot of *non*-pretense behavior as pretense. Consider, for instance, a mother driving her 36-month-old to nursery care. Behavioral accounts appear to predict

that the child will interpret the mother as *pretending that she's driving*; for, she surely behaves *as would be appropriate if she were driving*. Cases of this sort are very easy to multiply. Generally speaking, if *s*'s behaving *as would be appropriate if p* is a fundamental criterion for judging that *s* pretends that *p*, then nearly any behavior will ultimately count as pretending one thing or another. Hence, it is incumbent upon the behavioralist to say something more about how the pretense detector discriminates earnest *as if p* behavior from *as if p* behavior in pretense.

On the other hand, the behavioralist account seems overly *exclusive* with respect to pretense. For, the prediction (again, *prima facie*) is that the detector will tend to exclude instances which should clearly fall within the extension of pretense. This would appear to be the case, especially, in pretenses involving object-substitution (i.e. the use of props), and in those involving the use of attributed sound effects.

Consider, for example, a father pushing a pencil along the surface of a table while uttering 'Vroom!' Dad pretends *that the pencil is an automobile*, and even very young pretenders apparently have no trouble identifying the behavior as such. But Dad doesn't – *strictly speaking* – behave in a manner which *would be appropriate, if the pencil were an automobile*. Normally Dad wouldn't be making the 'Vroom!' noises – it would be the car making a (somewhat) similar noise. And normally Dad wouldn't be pushing a (tiny) automobile effortlessly, and with a single hand, on top of a table. In ordinary circumstances, Dad would actually be inside the thing. On reflection, in such cases it is not very clear at all how we should understand the relation *s acts as would be appropriate*

if p. Yet children have no problem recognizing and engaging in object substitution and sound effect pretense of this kind.

In the first place, Langland-Hassan counters, we needn't worry that the behavioralist pretense detector should radically *over categorize* pretense (i.e. that it should be too inclusive). For, by hypothesis, the detection system is sensitive to *both* cues indicating that a pretender acts *as would be appropriate if p*, *and* to cues that the pretender acts *as would be appropriate if not-p*. And for their part, the *as if not-p* cues are especially critical, as they include the stereotypical *as if not-p* cues which tend to accompany pretense generally: for example, the putative pretender acts in exaggerated and out-of-ordinary ways.[20]

Now, for example, when Mother is driving to nursery care, the situational cues surely indicate that Mother is behaving *as would be appropriate if she were driving*. But this is more or less *all* they indicate. There are, in particular, no salient and conspicuous clues that Mother is behaving *as would be appropriate if she were not driving to nursery care*. She's not, for example, focusing her attention on the child, looking at her with exaggerated facial expressions and making grandiose gestures, and so on. She's not, moreover, sitting on the floor of the playroom imploring, 'Let's go driving now!', and then singing 'The Wheels on the Bus' while turning an invisible steering wheel.

[20] Usually, behavioralists about pretense don't go much further than this in specifying what the relevant cues are. Suffice it to say they're cues like winking, singing, making exaggerated facial expressions and gestures, and so on.

Thus, in addition to acting *as would be appropriate if p*, Mother simply does not act as if playing any sort of game. Together, the abundance of salient (typical) *as if p* cues, combined with a lack of *as if not-p* cues, to include pretense-game-specific *as if not-p* cues – e.g. saying, 'Let's go driving now!' in an exaggerated and sing-songy voice – is supposed to have the combined (non)-effect of *not* activating the pretense detector. Thus, the charge of over-inclusivity on the part of metarepresentationalists is supposed to be unfounded.

Now return to the object substitution/sound effect cases. Why don't the ample – sometimes bizarre – *as if not-p* cues in such cases lead to a failure of detection? The activation of the pretense detector in such cases is supposed to follow attention to the general manner cues which indicate that a pretense is in effect. Here, for instance, unlike the case of Mother's driving, Dad surely is acting playful, silly, and out of the bounds of ordinary circumstance – to include focusing attention on the child, making exaggerated utterances and gestures, etc. It's evident from observable evidence that a game is in effect. Moreover, the behavioralist story goes, the pretense detector attends to salient analogical possibilities surrounding the relevant object/prop. Namely, attention is given to the fact that the object takes on property resemblances to some other object, with the specifics being provided by the manner cues available.

Thus, for instance (Langland-Hassan, 2011, p. 23; my emphasis):

> ‘the child recognizes that *the father is trying to make the pencil saliently car-like*…manner cues both direct her attention to the pencil, and allow her to recognize that he is starting a pretense game with respect to the pencil.’

In sum, detection in such cases proceeds as follows: Dad behaves, conspicuously, *as would be appropriate if the pencil were an automobile*. In particular: Dad makes the pencil saliently car-like. He does so by ‘driving’ it around, by saying ‘Vroom!,’ and so on. Moreover, Dad behaves, conspicuously, *as would be appropriate if the pencil were not an automobile*. He’s not sitting in it, he’s pushing it without effort, and he exhibits an exaggerated and playful manner that does not ordinarily accompany either Dad’s pencil-directed or automobile-directed behavior. Hence, the neo-behavioralist story goes, the cumulative effect of Dad’s manner is to activate the pretense detector.

In all this, the neo-behavioralist maintains, there is no reference on the part of the detector to the psychological states subserving Dad’s playful manner.

This tour of the neo-behavioralist account of pretense recognition leaves us with questions, to be sure. It’s not clear at all whether and how we ought to specify which behaviors and manner cues, either in kind or in specific cases, could be said to reliably activate a heuristic which determines that some appropriate ratio of *as if p* to *conspicuous as if not-p* has been reached, thus indicating that the concept GAME should be applied. I’ve been known to drive, sing, and eat a sandwich, all more or less simultaneously – but

my son doesn't seem to think that this combination *as if driving* and *as if not-driving* cues indicates that I'm pretending.

But notice that questions of behavioral evidence for pretense point to challenges for *everyone*, and it would be unfair to direct the challenge specifically at the behavioralist camp. Even the metacognitive account of pretense detection needs to tell *some* kind of story about *some* sort of mechanism for discriminating the observable evidence of pretense. Which behaviors and manner cues are supposed to lead to the activation of the *Decoupler* – to include the crucial inference that *s* pretends *(qua imagines*) *that p*?[21] In effect, even the metarepresentationalist must fall back on *some* representation of the outward and observable face of pretense games in order to explain why pretenders come to attribute mental states (of the relevant sort) in the right kinds of situations (those where the relevant cues are manifest).

Thus, I propose to focus in on this neo-behavioralist notion of a pretense game, in order to more fully assess whether it accommodates the data within its own theoretically determined constraints.

[21] Recall that even Leslie acknowledged the need for some additional cognitive machinery, which functions to link decoupled, imaginational representations to a pretender *via* the proposed (meta)representation of the proprietary informational relation underpinning pretense.

2. Why the Behavioralist Pretense Detector (Still) Doesn't Work

The following seems to be an accurate summary of the behavioralistic process which is supposed to lead to the identification of pretense:

(i) The detector judges that an agent *acts as would be appropriate if p* (in salient and conspicuous respects).

(ii) The detector judges that an agent *acts as would be appropriate if not-p* (in salient and conspicuous respects, especially to include the recognition of cues that some agent is playing a proprietary kind of game).

(iii) The detector judges that the agent makes some object (and/or state of affairs) saliently like some other sort of object (and/or state of affairs).

The last judgment, I think, could be paraphrased as 'The detector judges that the agent plays a game in which s/he pretends that *p*', but we needn't worry about the terminological details on surrounding this point.[22]

Now, to begin to see where the trouble arises for this account, start with a question: Could the behavioral cues tracked – in and of themselves – suffice as reliable indicators

[22]Notice that we are not limiting the discussion to cases of pretense involving object substitution. The same mechanism is supposed to be at work in ascertaining that *Dad pretends that we're at a tea party*, by detection of the set of observable facts amounting to, *Dad is making himself behave in saliently at-a-tea-party-like ways* (Langland-Hassan, 2011, p. 22-3).

of the kind of property the detector is supposed to track – i.e. pretense scenarios with particular contents? More to the point: Can it do so without recourse to a psychological concept of pretense (or to psychological concepts, generally)?

For his part, Langland-Hassan (2011, p. 13) acknowledges that one may well wonder whether imputing the described behavioralist concept GAME tacitly imputes a mentalistic understanding of pretense, *a la* the metacognitive account from Leslie:

> 'An obvious question is whether attributing to a child an understanding that a pretense "game" has begun secretly imputes to her an understanding of mental states. I think it does not…'

Further on (p. 21), he elaborates:

> 'In one sense, we certainly have ascribed the child the concept PRETEND, to the extent that being able to detect and play such games constitutes understanding pretense. In this (behavioral) sense of 'pretend', the child fully understands…pretending…. The important point is that we have not thereby given the child the concept of a mental state. Rather, we have given the child the concept of a kind of *game*, the recognition and playing of which does not require an understanding of mental states…'

Now, as I noted above, all parties require a detection mechanism for pretense games which is sensitive to the outward appearance of pretense. But what have we (*really*) attributed to the mechanism when we attribute to it this capacity? On reflection, all we've given the mechanism is the capacity to discriminate *a general category of social context*. In effect, we've awarded it the capacity to discriminate a proprietary set of behavioral-contextual conditions, in virtue of the detection of which it might be reliably caused to apply the concept GAME.

However, this is to say nothing about how the detector comes to a sensitivity to the *particular contents of these particular games*. Thus, we need to distinguish, on one hand, the detection that a pretense game is in effect, from, on the other hand, the detection that the game involved dictates that the players imagine particular contents. The neo-behavioralist account runs this context/content distinction together.

Actually, *both* the metacognitive and the behavioral accounts have tended to run the pretense context/content distinction together. But in the former case, there's good reason for this. Metacognitivists have traditionally promoted a view according to which pretense contexts are identified, in large part, by the identification of the contents of the special attitudes of pretenders. Since this psychological resource is not available to the neo-behavioralist, we now come to clearer picture of the challenge for the behavioralist approach generally: Could a pretense *context* detector – i.e. the behavioralist pretense game detector – do the *whole* job – to include the capturing of content? If so, how?

From what I can tell, the answer is supposed to be that the detector captures both context and content, and that it does so solely by tracking manner cues. The manner cues, for their part are supposed to indicate both that a pretense is in effect, and that the content of the pretense is *p*. The specification of *p* proceeds *via* tracking that the relevant agent *makes* some object (or state of affairs) *x* saliently like some other object (or state of affairs) *y*.

We're now in position to articulate the general form of a residual challenge for behavioralists: It is to produce a robust and non-circular behavioralist notion of '*s* makes *x* saliently *y*-like'. Unfortunately, there's a deep ambiguity in the crucial notion of 'making' that looks to make the challenge insurmountable. There surely is a robustly behavioralist notion of making – but it's obviously not the notion required for pretense detection. On the other hand, there's a notion of 'making' that's *very* well suited to accounts of pretense detection – but it's thoroughly psychologistic in nature. There doesn't seem to be any intermediate notion suited to the neo-behavioralist account of pretense detection.

Start with an uncontentiously behavioralist notion of making *x* like *y*. Call it the *transformational* notion of '*s* makes *x* like *y*':

(1) *s* transforms *x* in saliently *y*-like ways

According to this sense, we suppose *s* to literally alter the physical properties of the observable object *x*. In doing so, we suppose, *s* changes *x*'s physical properties, with the result that *x* is more like *y* after the transformation than it was before. In the pencil/automobile case, this would involve, for instance, pinning wheels to the sides of the pencil, drawing doors on the pencil, attaching a 'Vroom!-sound making device to the pencil, and so on. The effect is the production of a kind of physical replica of some object (or kind of object), *y*, out of some distinct object *x* by the literal transformation of *x*'s physical properties.

This sense of 'making' is a robustly behavioralist one. The trouble is that it's now here near the notion required for the description of a reliable pretense detection mechanism. The obvious prediction is that it would get all kinds of case wrong, and specifically that it would classify all manner of non-pretense, making-behavior as pretense. The mechanism should classify, for instance, the painting of a small plastic model of the R.M.S. Titanic as pretending that the model is the Titanic. For, in painting the model, I'd be transforming it saliently Titanic-like ways.

These problem cases multiply quickly. If I were to draw a map for you from my house to the nearest gas station, I'd be trying to make the lines on the paper saliently Brattleboro-like. But I surely wouldn't be pretending that the map was Brattleboro. Nevertheless, a pretense detector operating under the governance of a transformational notion of 'making' would classify these cases as pretense.

Langland-Hassan apparently recognizes the ambiguity in 'making', and tries to avoid the sort of worry I'm raising here by re-asserting the role of stereotypical manner cues in the recognition and inciting of pretense behavior:

> '…one can hold that the detection of manner cues *together* with detecting that *x* has been made saliently *y*-like is what the child uses to discriminate cases of pretending that *x* is *y* from cases of merely making *x* saliently *y*-like…[R]ecognizing that someone is pretending that *x* is *y* can be accomplished by recognizing that one is making *x* saliently *y*-like while engaging in some of a familiar cluster of manner cues, some of which focus attention on the subject matter of the pretense.'

But we cannot avoid the conclusion that a *transformational-making-tracker* would not be a very good pretense tracker, simply by requiring that stereotypical cues for pretense accompany instances of transformational making.

Let's grant, just for the moment, that the *transformational-making-tracker* only (usually) detects pretense when transformational-making accompanies stereotypical manner cues for pretense. Perhaps there are cases where the two overlap, and where such a detector would get it right. Such a detector would still be terrible at classifying pretense. For, the majority (probably the *vast* majority) of paradigmatic cases of pretense don't involve transformational making at all.

The typical case is rather more like the pencil/automobile case, where Dad says 'Vroom!', *thereby* (presumably) making the pencil saliently automobile-like. But such cases need not, and frequently do not, involve the transformation of the physical properties of the objects involved at all. In this case, Dad's the one making the noise. So what is the relevant sense of 'making' in the paradigmatic, object substitution pretense?

Here is a different, but similarly natural reading of 'making saliently like'. Call it the *property attribution* notion of '*s* makes *x* like *y*':

(2) *s* attributes properties to *x*, thus indicating that *x* is saliently *y*-like.

At first blush, this doesn't look like it's going to be much help to the behavioralist. First off, as it appears here 'attributes' is also multiply ambiguous.[23] Then, the most obvious candidate senses of the term each seem to be thoroughly psychologistic in their reference.

To a first approximation, for instance, 'attributes' might here mean something like 'judges'. For instance, to say that Dad attributes the sound 'Vroom!' to the pencil seems to be to suggest something like Dad judges, of the pencil, that it makes the sound. Of course, I doubt anyone – behavioralist or otherwise – would want to stand by such a reading. For, while 'judges' is also ambiguous, the most natural interpretation is a sense in which it involves the attribution of a belief. In the example, it would be the attribution

[23] For that matter, so is 'indicating'. However, it will do for present purposes to focus attention only on 'attributes'.

of the belief with the content that the pencil has a property that makes it saliently automobile-like. Again, this surely can't be the relevant sense of 'attribution'.

Another possibility is that 'attributes' means 'judges' in the sense of 'judges in imagination'. In effect, that it means something more like 'makes believe'. Thus, in our example, the crucial piece that the detector tracks would be the fact that *Dad makes believe that the pencil exhibits 'Vroom!'-noises*. Since the observable evidence sorely underdetermines the content the detector needs to grasp, this is a natural and tempting way to understand the notion of making subserving pretense behavior and the recognition thereof. This notion does not, unlike the transformational notion, implausibly require that the objects and persons involved in a pretense be literally changed in any way.

Of course, none of this will do for the behavioralist. For, now in all but name, the claim is that the way the detector captures (in particular) the content of the pretense is by noticing that Dad *pretends* – in a psychologistic sense of 'pretends' – that the pencil has automobile-like properties. The arguments here are surely not decisive. Nevertheless, barring some consistently behavioralist alternative to the transformational and the property attribution senses of 'making', we're led to the conclusion that pretense recognition is indeed metarepresentational. No such alternative appears to be forthcoming.

3. *Conclusion*

Contra the way Langland-Hassan sets up the dialectic, the real trouble for the behavioral account is not really in ensuring that the pretense detector reliably discriminates cases of *pretending that x is y* from cases in which *s* makes *x* saliently *y*-like (in the transformational sense of 'making'). The real trouble, as I've argued, is in giving a truly behavioral account of 'making *x* saliently *y*-like', in the first place.

The account canvassed here does not provide such an account, and the result is an unhappy dilemma for behavioralism about pretense generally: Either the relevant sense of 'makes *x* saliently *y*-like' is supposed to be the transformational notion in (1), or it is supposed to be the property attribution notion in (2). Say it's (1). Then we have our behavioralist account, but it doesn't capture the right extension of behavior. So, it must be some suitably disambiguated version of (2). But upon further analysis, it looks as though the notion of making behind (2) that we wind up with – while it promises to capture the right extension of behavior – is too psychologistic. It amounts, in effect, to something more like the sort of making involved in 'making believe'. The kind of 'behavioralism' that we'd wind up with would be little more than a notational variant on the metacognitive view.

It's perhaps worth noting, in closing, that my arguments here point to a very general, and historically recurring pattern in the dialectic between cognitivist and behaviorist styles of explanation in psychology.

Generally speaking, behaviorists aim for economy of explanation, to include (especially) minimizing commitments to unobservable theoretical posits. They opt, instead, to use *prima facie* non-psychologistic terms like 'conditioning,' 'reinforcement,' and 'association.' But terminology, as we've seen, can be very misleading, and upon analysis it seems likely that even these most fundamental behaviorist notions implicitly involve reference to internal psychological states. It's unclear, for instance, whether terms like 'reinforcement' can be adequately defined without reference to what a creature wants, to what it likes, or to the feelings evoked in it by its rewards and punishments (Chomsky, 1959; Levin, 2013). At least, no behaviorist has convincingly argued otherwise.[24]

I submit that the discussion of pretense cognition exhibits a manifestation of this same dialectic. Consistent with the pattern, further analysis of the language of the behavioral account has bolstered the representationally rich, metacognitivist account. Ultimately, it does in fact seem that genuine pretense-game recognition, to include the detection of both context and content, requires attributing a stronger link between agents and the contents of the pretense than a thorough-going behavioralist account allows.

[24] There's a similar parallel to observe in the philosophy of language. So-called analytical, or logical behaviorists (Ryle 1949) try to analyze the reference of terms – e.g. 'belief' – solely in behavioristic terms. Accordingly, the meaning of sentences like '*s* believes that it's snowing' is supposed to be given by a set of behavioral dispositions for *s*. For instance, the meaning of this sentence would be something like *s* is disposed to put on his boots, to put on his hat and gloves, and so on. But notice that this kind of semantic analysis is plausible only given auxiliary assumptions about *s*'s psychological states – for instance that *s wants* to stay warm outside, or that *s aims* at *comfort*, and so on.

Chapter 4: A Pluralistic Account of the Propositional Imagination

'Given the increasing importance that imagination has been assigned…it is important to try to get clear on what this mental activity is…Thus we have [a] question…[I]s there such a thing as the phenomenon of imagining, i.e., is there a single mental activity that can do all the explanatory work that has been assigned to imagination?' (A. Kind, 'The Heterogeneity of the Imagination,' 2011)

In this final chapter I begin to sketch a framework for an alternative architecture of imagination. The architecture is novel in that it openly eschews the long-standing *desideratum* that an architecture of the imagination should provide a unifying account of imagination, built around a singular variety of representational vehicle. Instead, I offer a pluralistic account of propositional imagining, in which I suggest that the term 'propositional imagination' tracks multiple, distinct psychological phenomena.

Pluralism about the imagination is already beginning to be explored in the philosophical literature. As I've observed throughout, and as Brian O'Shaugnessy (2000, 339-340) has pointed out elsewhere, there is surely a marked tendency to take it *for*

granted that 'there exists something that is the Imagination' or that 'there is some one thing that is the phenomenon of Imagining.' But a less popular possibility, recently proposed by Amy Kind (2011) is that we're ultimately going to have to give up on the quest for a unified account. In short, to invoke Kind's terminology, it's beginning to look as though the imagination is heterogeneous. If so, then accounts traditionally cast in terms of a singular cognitive capacity, subserved by a special kind of mental representation, need to be recast in terms of the plurality of psychological capacities, representations, and attitudes underpinning imagination.

Now, Kind's purpose is general; it's essentially to support the broad claim that imagination splits into a heterogeneity of psychological phenomena. Moreover, it's evident from her arguments that the relevant sense of 'heterogeneity of psychological phenomena' is 'diversity of psychological attitudes' – though I do not mean to discuss the details of her reasoning here. What I aim to do, rather, is to (as it were) pick up where the style of argument Kind offers leaves off, and to begin to sketch a taxonomy of the heterogeneous propositional imagination.

While it does seem like the phenomena I'll describe have commonalities, I mean to remain neutral (for the time being) on the question of whether we should cast propositional imagination as a superordinate natural kind, or whether we ought to view the relevant phenomena as wholly distinct in kind.

Before setting out, I can also offer a pair of very general motivations for entertaining a pluralistic alternative to metacognitivism, new cognitivism, and the single attitude accounts. The first motivation stems from empirical considerations I've raised along the

way, especially in my reviews of the new cognitivist and single attitude approaches to imagination. The other motivation stems from a point of methodology.

First: In spite of insights into the nature of imagination gained from the metacognitive, new cognitive, and single attitude proposals, I've argued that each account has trouble explaining various aspects of the outward faces of imagination and pretense, on one hand, and/or their subserving mechanisms, on the other. Many of the difficulties raised manifest, I wish to propose, a continuation of the persistent tendency to unify propositional imagining under the auspices of a single psychological category. I'll have occasion to revisit some of these difficulties in more detail, and to show how they're avoided by adopting imaginational pluralism, below.

Second: I acknowledge that one might think that adopting a pluralistic stance is undesirable – for, we seem to lose both economy and parsimony, on one hand, and to lose the explanatory virtue of unification on the other. Hence, isn't adopting a pluralism about imagination ultimately *too* costly?

On reflection, I don't think pluralism is very costly at all in this case. At least, it's not costly enough to render the strategy *prima facie* implausible. For, the empirical considerations in favor of pluralism, as we'll see, are compelling. Moreover, notice that *everyone* has already granted that (at least) the folk notion of imagination splits in its reference. The term 'propositional imagination' is a term of art, coined to make reference to what's viewed as the capacity to imagine propositional content, as distinct from the capacity to token mental imagery.

One way to understand what I'm doing here is that I'm pushing this trend – the trend of developing a finer-grained notion of imagination – a step further. My claim, essentially is that, like the folk term 'imagination', the technical term 'propositional imagination' splits further in its reference. Propositional imagining, in effect, is underpinned by multiple, distinct phenomena, none of which is independently capable of accommodating all of the relevant data.

Now, to begin to get purchase on the varieties of imaginational cognition, I will to single out and discuss three representational vehicles of propositional imagining. I do not mean to suggest that these three exhaust the cognitive underpinnings of imagination. To the contrary, I'm happy to incorporate more vehicles, if and where necessary.

The three vehicles of imagination I'll canvass here are, in preview, counterfactual elaboration over the content that *p*, (merely) entertaining that *p*, and fictionalized attitudes toward the content *p* – i.e. attitudes toward known fictions and pretenses. Hence, on a final point of motivation, there's a sense in which the kind of pluralistic account I have in mind is, in fact, a parsimonious alternative. For, the phenomena which I suggest comprise the primary vehicles of imagination are admitted (or should be admitted) more or less by everybody to be (*i*) indispensable components of our cognitive architecture, while (*ii*) being broadly imaginational in nature.

Thus, on the grounds of empirical considerations, methodology, and parsimony, I maintain that pluralism is not too costly of a strategy to adopt in the case of propositional imagination.

1. Counterfactual Elaboration: Believing that if p were, then q would be

Start with an overview of propositional imagination *qua* counterfactual elaboration. The single attitude hypothesis (*SAH*), as developed by Peter Langland-Hassan (2011) proposes to assimilate the mechanisms of propositional imagination to those subserving the counterfactual elaboration over the consequences of particular propositions.

The idea has a lot of *prima facie* plausibility. Who would ever doubt, to begin with, that we very regularly consider the kinds of 'what if *p'* questions which, as the *SAH* story goes, drive the propositional imagination? On this view, imagining that *p* is just the sort of thing we do when we ask ourselves, for instance, *what would the world be like if I won the lottery tomorrow*? The sort of process the single attitude account is developed around is a familiar component of our cognitive architecture.

It's also uncontroversial that counterfactual reasoning bears crucial mechanistic, semantic, and conceptual relations to imagination (Currie and Ravenscroft, 2002; Evans and Over, 2004; Byrne, 2005). Broadly speaking, psychologists and philosophers have usually tended to think about counterfactual reasoning as a kind of imagining, or as a kind of imaginational activity or process.

Nevertheless, I've argued that counterfactual elaboration, as an imaginative process, cannot accommodate very large tracts of data usually associated with propositional imagination. Namely it's implausible to hold that counterfactual reasoning is the primary mechanism of fiction consumption (or production, for that matter; see chapter 2).

Counterfactual elaboration seems likewise insufficient as a mechanism of pretense detection (chapter 3).

The moral, of course, is not that we should abandon the idea that counterfactual elaboration counts as propositional imagining, because it fails as a vehicle which enables us to give an elegant and unifying account of the vehicles of imagination. To the contrary, we need to expand the varieties of vehicle we incorporate into the architecture.

2. *Bare Imagining: Entertaining that p*

The second kind of imaginational vehicle I have in mind are the kind of '*bare imaginings*' I referred to briefly in chapter 2. Bare imaginings involve the raw, or mere, representation of a content in the alethically and epistemically neutral mood frequently associated with imagination.

The central strategy of the new cognitive theory (chapter 2), I take it, was essentially to single out bare imaginings, and to apply the functionalist strategy in arguing that – given a *prima facie* plausible set of distinctive causal features for bare imaginings – the vehicles of propositional might be assimilated to, or rather identified with, bare imaginings.

I argued at length in chapter 2 that the new cognitive theory is unsuited to capture a distinctive causal role of imagination, *qua* unified, *sui generis* psychological kind, by pressing states of bare imagination into service. For, new cognitivism is at its best when

explaining the similarities between belief and imagination – *ex hypothesi* there're no compositional differences across believing and (barely) imagining that *p*; thus, the story goes, they bear many causal similarities. It turns out, however, to be surprisingly difficult to tell the other half of the story – (*i*) what properties of beliefs and imaginings account for the (putative) fact that they should systematically different causal profiles? And (*ii*) after all, do beliefs and imaginings have systematically different causal profiles? In the end, the summary idea that we seem to get from the new cognitive account is that imagination is really *a lot* like belief, it just doesn't have a few of the effects that belief does. For example, it plays no motivational role (Nichols and Stich, 2003)

But I doubt, anyway, that bare imaginings were ever the right sort of vehicles to assimilate to the kinds of causally robust imagining we see in counterfactual elaboration and (we'll see later) in fictionalized attitude imagining.

A more promising way to understand the difference between bare imagination and mere entertaining, and their (in principle) compositionally isomorphic counterparts in belief, is by attending to the primitive and basic cognitive nature of the latter. As Scott Soames (draft) has recently observed, for example:

> 'Entertaining a proposition is the most basic attitude we bear to it. It is the attitude on which the others are, in one way or another, based.'

Variations on the idea that entertaining is somehow basic and semantically prior to the other attitudes, including belief, is fairly common. There's a tradition, for instance,

according to which imagination is the mental analogue of (mere) utterance, where belief is the mental analogue of assertion (Scruton, 1974; McGinn, 2004). To the extent that the analogy is illuminating, it suggests that the entertaining states of bare imagination are prerequisites for having propositional attitudes generally. For, believing that *p*, desiring that *p*, hoping that *p*, fearing that *p*, and so on all imply entertaining that *p*.

The utterance/imagination analogy is suggestive of another way in which bare imagination seems to be primitive and basic. It suggests that the primitiveness of bare imagining might be intimately tied up with some other, very general and basic, properties of cognition. For instance, it suggests that (barely) imagining that *p* is a manifestation of our general capacity to *systematically* and *productively* manipulate and transform mental contents generally, and more or less at will. Mental representations, *prima facie*, mirror sentential representations in that there're few constraints on the amount, or nature of conceptual and propositional components we can voluntarily combine in a complex mental representation. Try a quick series of cognitive experiments, entertaining the following contents in bare imagination:

(i) *that there is a highest prime number*

(ii) *that if there are more than twelve million people in New York City, then Washington never crossed the Delaware*

(iii) *that the red hat sailed in the wind over the rolling green hill after dusk but before dawn and near to the 1st of the year*

Mere entertainings, such as these, seem straightforwardly to be exercises of a simultaneously productive and imaginational capacity. Bare imaginings seem equally implicated by cases of the systematic manipulation of contents, as illustrated by the following kind of sequence:

(iv) *that Bill loves Hilary*

(v) *that Hilary loves Bill*

The idea is just that if you can entertain (iv) there's no in principle reason why you shouldn't be able to entertain (v). And in practice, it indeed seems easy to token both (iv) and (v). But doing so doesn't seem to (necessarily) involve one in doing anything other than (barely) representing the content. I needn't have any other commitments regarding Bill and Hilary, for example that might lead to the spontaneous production of further inferences or the production of associated imagery.

The point, to be clear, is not the grandiose one that the vehicles of bare imagination comprise the cognitive underpinnings of our capacity for productive and systematic thought. It's more likely the other way around: One consequence of harboring a capacity for productive and systematic thought is possessing a further capacity to token a peculiar kind of representational vehicle – i.e. I'm calling them bare imaginings, or mere entertainings – tokens of which often wind up playing no substantive cognitive role at all.

3. *Fictionalized Attitudes: Representing that, in the fiction, that p*

The last kind of imaginational vehicle I wish to discuss are fictionalized attitudes.

To a first approximation, the imagination frequently features in the tokening of propositional attitudes, the contents of which we *explicitly* regard as non-actual or fictional. Taking my cue from the metacognitivist camp, I maintain that the relevant cases involve tokening propositions which fall within the scope of a concept, or perhaps a scope operator, which make explicit reference to a fictional, imaginary, or pretend world. The relevant propositions thus derive a fictionalized valence from the wider-scope contents. Thus, in contrast to bare imaginings (and to counterfactual elaboration, for that matter), some imaginings are – as it were – self-aware of their imaginational status.

As with counterfactual elaboration and bare imaginings, positing attitudes of this kind shouldn't be very contentious. To be sure, there's been a fair amount of debate over whether very young children are capable of producing fictionalized attitudes, in particular, in the context of pretense. I've already argued (chapter 3) that we have good reason to believe that even very young pretenders understand that pretense involves imagination. We thus ought to side with the metacognitivist on the issue of pretense recognition; for, we have every reason to think believe that the only reliable means of pretense detection proceeds *via* the production and storage of states with metarepresentational contents, e.g. S PRETENDS THAT P, or IN THIS PRETENSE, S BELIEVES THAT P, and so on.

After all, however we wind up cashing out the content of the concept PRETEND, everyone is in agreement that the concept must refer to a context in which some (as it were) reality-cancelling assumptions are in play. In contrast with case of child's play, I take it that the claim that psychological states of this general variety are a part of the cognitive repertoire of *adults* is not controversial in the least. I doubt anyone would deny, for instance, that adults recognize pretense by assuming that the observed subject pretends that *p* – i.e. that they imagine that *p*, or otherwise bear some special informational relation to the content *p*.

On a final point about fictionalized attitudes, there doesn't seem to be any reason to restrict the contents of what counts as a fictionalized attitude to metacognitive representations. We seem to have beliefs about the contents of fictions which, though they make no reference to the psychological states of any individual, ought to count as fictionalized attitudes. Whenever we have a belief with contents of the form, *that, in the fiction, that p*, for instance, we represent *p* in the (as it were) detached or alethically neutral mood of imagination.

Clearly, cases of fictionalized attitude imagining will multiply very quickly. The category will incorporate any token attitude, to include beliefs, desires, hopes, fears, and so on – which makes explicit reference to some piece of fiction, some story, some movie or play, some pretense, or any other medium which we take to be fictional in nature.

Notice that there's very little motivation for insisting that fictionalized attitudes of kind described here do not *really* count as propositional imaginings. On one hand, given

the arguments from previous chapters, we simply cannot *assume* that there's some obvious candidate that we can assimilate to *the* vehicles of imagining that *p*. The state of the debate, as I've presented here, suggests otherwise.

Moreover, it bears emphasizing that it's rather uncontentious to assume that pretense, even in small children, involves the pretender in exercising imagination. But for all that's been said, this is just to say that pretending involves pretenders, on the one hand, in counterfactual elaboration (*a la SAH*). The single attitude account, we saw, goes a long way toward explaining what goes on cognitively in simple pretense games, without appealing to either the distinct attitudes of new cognitivism, to what I've called bare imaginings here, or to metacognitive imaginings of the kind proposed by Leslie and his collaborators.

On the other hand, I've argued, exercising imagination in pretense involves – perhaps in *addition* to counterfactual elaboration – metacognitive imaginings which serve (at least) in the detection of pretense by both children and adults.

Hence, given the range of what intuitively counts as involving an imaginational process, there simply doesn't seem to be anything theoretically fruitful to gain by reserving the term 'propositional imagining' for some subset of vehicle – say, bare imaginings – which is itself incapable of explaining the full range of what has traditionally fallen within the purview of the theory of propositional imagination.

4. Three Vehicles of Imagination at Work: Some Problem Cases (Revisited)

Having given rough and ready sketches of several vehicles of imagination, we're in position to get a provisional view on how the pluralistic approach can help to sort some of the puzzles I raised in earlier chapters for the available accounts.

4.1 Imagination and Pretense

An important empirical puzzle with pretense, as it has traditionally been presented (and as I've reviewed it here; chapter 3), is over whether very young pretenders need to have some kind of metacognitive capacity in order to detect pretense in their playmates. It's also an open question whether young pretenders rely on metacognitive representations in order to engage in pretense – e.g. in order to reliably effect the reality-cancelling presuppositions that apply to the contents of typical pretense games, on the other.

As I argued previously, and then reiterated above, the behavioralistic account of pretense recognition – which is incorporated into both the new cognitive and the single attitude accounts – is ultimately untenable. For, while everyone agrees that pretense detection requires some concept of pretense, there is no adequate, behavioralistic version of the concept PRETEND forthcoming. If I'm right, then fictionalized attitudes roughly of the sort originally posited by the metacognitivists about pretense – i.e. cognitive states of general form AGENT PRETENDS THAT P – are the most plausible candidates for the

primary mechanism involved in detection of pretense games. It's uncontroversial that these kinds of attitudes are relied upon by adults.

But if fictionalized attitudes are relied upon (even by very young pretenders) for the *detection* of pretense, then *a fortiori* they're available for the deliberation over and elaboration upon the contents of pretense games. Young pretenders, more or less as Leslie (1987) predicted, ought to be able to draw upon their store of background knowledge in the construction and tracking of pretend scenarios as complex as their knowledge and attention will allow. Psychologically mature adults should exhibit an expanded capacity for pretense, given their wider knowledge base and an augmented attention span.

Finally, we should expect the fictionalized attitudes of both young and old pretenders to play a motivational role in pretense. There's no *prima facie* reason to assume that fictionalized beliefs and fictionalized desires could not interact to produce pretense appropriate behavior. And, as the case for *SAH* convincingly lays out, counterfactual elaboration provides an additional path through motivation to game-appropriate action in pretend play. There may be a role for bare imaginings here, too, if we assume they're able to be taken as input by the counterfactual reasoning mechanisms, *a la* the query states of *SAH*, to introduce novel contents into a given pretense.

It bears mentioning that bare imaginings, in and of themselves, appear to be ill-suited to the purposes of driving pretense appropriate behavior. For, presumably no one wants to award bare imaginings – *qua mere entertainings* – any real motivational role.

Interestingly, the same point crops up in the guise of a challenge for the new cognitive account. Reluctant to grant a robust motivational role for their distinct attitude imaginings, Nichols and Stich (2003) appeal to a process of counterfactual elaboration. In short, as the distinct attitude imagining that *p* is written to the pretense box, a series of imagination-tracking counterfactual beliefs of the form *if were, then q would be*, unfold in the belief box. And it's the beliefs, but not the imaginings, which are supposed to drive the action in the pretense game.

The reason this particular issue arises, I maintain, is due to the new cognitivist attempt to take a very clear candidate vehicle of propositional imagination – i.e. bare imaginings – and to try to construct a unified theory around that vehicle. In general, as I've already argued, this strategy yields a number of problems related to the specification of a causal profile for bare imagination. Since, for instance, imaginings are supposed to be taken as input by affect and inference mechanisms, why are they supposed to be eschewed by the action production mechanisms? There's apparently no satisfying answer forthcoming from the new cognitivist camp.

At least tentatively, my characterization of bare imaginings avoids this kind of problem by eschewing the central idea behind the distinct attitude hypothesis (chapter 1). That is, by eschewing the idea that (bare) imaginings – or mere entertainings, if you like – play a robust functional role, characterizable by their systematic interactions with the suite of paradigmatic cognitive mechanisms.

It's also worth noticing, finally, that in his statement of *SAH*, Langland-Hassan (2011) goes to great lengths to eschew distinct attitude imaginings, even the stripped down version of bare imaginings I take on here, altogether. For, as he apparently saw things, admitting bare imaginings would be to give up the game against the new cognitive theory, since it would apparently be to admit a form of distinct attitude imagining.

Nevertheless, *SAH* does incorporate an argument against the incorporation of bare imaginings into the architecture, and it would be unfair not to at least give a cursory treatment of how it works. Let's call it the 'needless or useless' argument (Langland-Hassan, 2011, p. 14):

> '…representing that p – be it in the [pretense box], or anywhere else—will result in relevant inferences being made only if one already has beliefs about what is generally true of situations where p…. And if one already has those beliefs, there is no need to represent that p in order to retrieve them; a desire and intention to determine what would happen if p will suffice. If, on the other hand, one has no beliefs about what would likely happen if p…nothing will emerge as reasonable when one represents that p in the [pretense box] – for, by hypothesis, there is nothing in the [pretense box] other than p and the copied contents of the Belief Box. So, representing that p in the service of trying to determine what would happen if p is either needless or useless…'

I don't intend to provide a full analysis and evaluation of the argument. For all that I have to say on the matter, it may very well be true that bare imaginings are needless or useless when it comes to the processes of counterfactual inference. They may be so for nearly all cognitive purposes. This would be consistent with my characterization of bare imaginings. The trouble is that, despite the granted cogency of the argument, it does not convince. For, as I argued above in my characterization of the vehicles of entertaining that *p*, bare imagining is ostensibly something we do, whether it serves any causal purpose or not.[25]

The way I see it, the real motivation for this sort of move seems mostly to be to keep *SAH* economical, especially by keeping the theory unified around a singular kind of psychological vehicle – in this case, the productions of the belief box in counterfactual elaboration. As I argued above in introducing the pluralistic approach, these really are not very moving considerations.

4.2 Semantically Rich & Semantically Poor Imagining

I initially canvassed problems of semantically rich and poor imaginings as problems for the single attitude account. Rich (propositional) imagining occurs when we engage the detailed descriptions of fictional worlds given to us, for example, in novels, short stories,

[25] It bears noting, too, that despite the argument offered, it's far from clear that *SAH* really gets by without bear imaginings. The account presently lacks an adequate account of the interrogative query states that are supposed spark counterfactual elaboration. One natural way to think about the vehicles of these query states is that they're bare imaginings, taken as input by counterfactual inference mechanisms. It's unclear whether, even granting that they might be needless, bare imaginings put to this use would (really) be useless.

films, and (perhaps) theater productions. To cast the kind of imagining we do when engaging these media as counterfactual elaboration over propositional contents, I argued (chapter 2), is implausible.

It's worth observing, too, that the data surrounding rich imaginings are not obviously accommodated very well within the new cognitivist framework, either. If, as I've suggested, the kinds of so-called 'distinct attitude' imagining that the new cognitivists assimilate to propositional imagining generally turn out to be more akin to the bare imaginings I've canvassed above, then they don't seem to be the right sort of vehicle for the representation of the contents prescribed in rich fictional worlds. For, bare imaginings, for their part, are ordinarily divorced from the rich fabric of contents, imagery, and associations we usually associate with the relevant sort of fictional worlds.

None of this is to say that counterfactual elaboration and bare imagining don't play importantly, even typically, in our engagement with semantically rich fictions. But even taken together, I maintain, they could not tell the whole story. The resources provided would lead us, *prima facie*, to give the following kind of account. The fiction consumer is presented with a bunch of prescribed propositional contents, which are tracked and represented, presumably, in bare imagination. The counterfactual reasoning mechanisms, let's say, take bare imaginings as input, and generate beliefs about the consequences of the contents provided.

This sounds like a legitimately imaginative process, perhaps even of the sort the single attitude approach should incorporate in order to flesh out the *SAH* account of

inferential elaboration in pretense. But it doesn't sound much like what we do when we engross ourselves in a richly described fictional world. It sounds rather like we've begun to sketch an account of the mechanisms of supposition, hypothesizing, prediction, and assumption – all perhaps *additional* vehicles of propositional imagination (or perhaps some combination of additional vehicles), each of which I'm going to leave aside for present purposes.[26]

Our experience of fictions (and elaborate pretense) corroborates. Our engagement with fictions is generally more as of *make-believe*, and make-believe – I'm suggesting – is in many cases just the having of beliefs about what goes on in a fictional or pretend world. Notice that it seems uncontentious that we can have desires about such worlds, too. In which case, it looks as though all the ingredients are present for a very rich, potentially emotionally charged experience. The kind of experience we should predict, if it includes our having fictionalized attitudes toward the contents that obtain in the relevant world, is surely different from the kinds of experience that would be yielded by the other vehicles of imagination. We are not, on one hand, playing a game which involves us in behaving in the stereotypical ways associated with pretense, while deliberating over the counterfactual consequences of more or less randomly chosen or entrained propositions. On the other hand, we are not engaged in the merely entertaining of propositional contents in the semantically sparse and disconnected mood of bare imagination.

[26] Though it bears emphasizing that there's a strong current in the philosophical literature which upholds the distinction between the imagining involved in fiction engagement and imagining *qua* supposing or hypothesizing.

Thus, the fictionalized attitudes rehearsed above – to include metacognitive attitudes about the imaginational states of ourselves and others – seem especially well-suited to understanding fiction engagement. Let's say I believe, for example, that, *at the world described in whatever book I'm reading, my favorite character is about to be killed off*. Probably, I'll become anxious and/or sad. This kind of phenomenon is often put forward as a puzzle, since I'd probably be happy to assert that the character doesn't exist. But on one hand, and *prima facie*, it seems like, given my belief, and my presumptive desire that the character endure, we should predict that I'll become sad or anxious. *Prima facie* the fact that the character is fictional, and arguably does not exist, does not seem relevant. I don't deny that there's a logical puzzle here, but it's not obvious that there's a corresponding *psychological* puzzle, when understood this way.[27]

Notice, too, that the varieties of fictionalized attitudes may also multiply. Hence, a fictionalized belief toward a pretense may be less pregnant with respect to affect than a fictionalized belief toward the world described by a television show that's been on the air for multiple years. For that matter, we may simply have stronger (fictionalized) desires toward the latter than we do toward the former.

For their part, I needn't say very much about semantically impoverished imaginings. This was the sort of case like *Chicken Farmer*, rehearsed in chapter 2, which was supposed to be a kind of single proposition story with the singular content *that the*

[27] It is also up for grabs whether the assertion 'that the character does not exist', on my part, is consistent with my total set of beliefs. Peter van Inwagen (1977) has argued that we are committed to the existence of fictional objects, as they appear to be required by a broader theory of literature we ostensibly tend to espouse.

chicken farmer at breakfast. The natural thing to suppose is that imagining of this kind – or rather this grain or gradient – is typically subserved by the bare imaginings I described above.

On one reading, in fact, the characterization of mere entertainings I gave was really just a characterization of a vehicle for representing semantically sparse contents. It doesn't really matter whether we're considering cases that are supposed to be stories. It's just that, when we call a case like *Chicken Farmer* a *story*, it seems odd for not having more prescribed contents than it does. But if we're admitting bare imaginings into the architecture, they'll allow for the more or less effortless representation of sparse fictional worlds, by the same mechanism they enable for the representation of random propositions, and manifest our general capacities for productive and systematic thought.

In sum, there's every reason to believe that all three of the vehicles canvassed here are at work in our engagement of rich (and poor) fictional worlds. In combination with mental imagery, the deployment of multiple vehicles of imagination only adds to the rich and complex representation of a robustly described fictional world. Given our capacity for bare imagination, the semantically sparse cases do not seem especially puzzling. Nevertheless, it's our fictionalized attitudes that really connect us, in personal and affect-producing ways, to fictional contents.

4.3 Bizarre Fictions & Flights of Fancy

The relevant cases here, recall, were *Nebraska Jack* and *Alien Invasion*. In the first case, we were asked to imagine that Jack at dozens of donuts and gallons of coffee for breakfast; in the second, that an alien soldier appears, along with his army, and offers me a candy bar and free dry cleaning.

Bizarre fictions and flights of imaginative fancy like these likewise looked like they would not be readily accommodated by the imaginational process of counterfactual reasoning described by *SAH*. It's hard to see, in particular, how a normatively guided process of belief production could plausibly subserve a process which generates representations of radically out-of-the-ordinary consequences and associations, as in *Nebraska Jack* and *Alien Invasion*.

On reflection, it's unclear how the new cognitive approach would deal with such cases, too. While the new cognitivists do not assimilate imagining that *p* to a process of counterfactual elaboration, they nevertheless uphold a strong relation between imaginational elaboration and background knowledge. Imaginational elaboration in the pretense box is supposed to proceed spontaneously, and in accord with default assumptions about reality. Hence, when I upend my (pretend) tea cup, I'll spontaneously be led to infer that the cup is empty. *Prima facie*, my charge that counterfactual elaboration is ill-suited to underpin bizarre imaginings and flights of fancy will apply, *mutatis mutandis*, to the new cognitive theory.

A natural thought – one I think we should remain open to – is the idea that, in some cases, bare imagining is at work in the representation of bizarre fictional contents. Part of what's distinctive about bare imagination is not merely the comparative semantic poverty of the imaginings it allows us to entertain, but also that it's a representational vehicle that allows us, in principle, to represent more or less whatever we choose. Hence, the contents of bare imaginings can be as bizarre or as fanciful as you like.

Nonetheless, I doubt that bare imaginings are the right vehicles for the sort of work behind *all* cases of bizarre imagining. In particular, there two sort of case I think bare imaginings won't accommodate.

The first is sort of case is just a variant of rich imagination cases. I argued above that fictionalized attitudes are the best available candidate vehicle for the work of capturing rich fictions, part of the representation of which involves the tracking that the relevant states of affairs occur at a particular alternative world (or at least in a world outside of our own). Rich, yet bizarre imaginings, yield the same kinds of considerations. These kinds of cases share the important feature with rich fictions that the attitudes that we bear toward the contents is akin to what we would think of as make-believe. And this is to say that, to the extent that we become engrossed and personally invested in a fiction with bizarre contents, it's presumably not in virtue of the fact that we *barely imagine* its contents. It seems rather to be in virtue of our directing, *inter alia*, fictionalized beliefs and desires at the relevant world.

Notice that attributing the relevant work to fictionalized attitudes dissolves the puzzles raised over rich and bizarre fictions for *SAH*, too. It's unclear how the representation of such contents – e.g. *Alien Invasion* – could be produced by a process of counterfactual elaboration. But it seems uncontroversial that such contents could be the objects of, for instance, fictionalized beliefs and desires. Deferring to these vehicles would allow us to say, for instance, that consuming very reality-deviant fictions proceeds by our representation of what we take to be the case in the world described. And the production of flights of fancy could proceed in accord with whatever we desire to obtain, at the relevant world – to include (perhaps) a sensitivity to genre-guiding desires. If I'm in the mood to produce a kind of alien farce, for example, I may come up with something like *Alien Invasion*.

None of this is to rule out a role for either bare imaginings or counterfactual elaboration in the consumption and production of bizarre imaginational contents. The point is just that neither of those vehicles are plausible as the primary vehicles in the kinds of cases I'm concerned with here.

5. Conclusion

I've argued that we should adopt a pluralistic approach to understanding the propositional imagination. The range of phenomena that need to be accommodated by a theory of

propositional imagination is far reaching. Here I've only concerned myself with a comparatively small cross-section of imaginational phenomena. I don't pretend to have delimited all the diverse vehicles subserving the phenomena associated with propositional imagination. Instead, I have suggested that there are at least three distinct varieties of psychological vehicle – to include distinct varieties of representations and distinct processes – each subserving an aspect of the phenomena which have been (perhaps misleadingly) consolidated under the auspices of the technical term 'propositional imagination'.

Moreover, it bears emphasizing that the scope of the theory of propositional imagination has yet to be adequately circumscribed. Just to gesture at a few important areas for further research – I've said nothing of the nature and mechanisms of supposition, assumption, hypothesizing, prediction, and conceiving. Neither did I touch on the intuitive relationship of the pluralistic vehicles of propositional imagination to the modal attitudes – i.e. beliefs regarding possibilities and necessities of different kinds.

I take it, given the sheer range of phenomena, that we should expect the vehicles of propositional imagination to multiply beyond the three basic vehicles canvassed here. For, there's every reason to predict that the pattern traced in the first part of this dissertation would continue. That is, it's plausible to expect that whenever there's an attempt to produce a unified account of propositional imagining in the service of explaining some comparatively small subset of data, difficulties in explaining other data

will inevitably arise. The case for pluralism, I take it, is made all the stronger for the apparent robustness of the pattern.

References

Boden, M., 2003: The Creative Mind: Myths and Mechanisms. London: Routledge.

Boyd, R., 1991: Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. *Philosophical Studies* 61: 127–148.

Bronte, C., 2010: Jane Eyre. New York: Harper Collins.

Byrne, R., 2005, The Rational Imagination. Cambridge: MIT Press.

Carruthers, P., 2006: The Architecture of the Mind: Massive Modularity and the Flexibility of Human Thought. Oxford: Clarendon Press.

Chomsky, N., 1959: "A Review of B. F. Skinner's 'Verbal Behavior'," *Language*, 35, 1: 26–58.

Currie, G. 1995a: Image and Mind: Film, Philosophy and Cognitive Science, Cambridge: Cambridge University Press.

——, 1995b: The Moral Psychology of Fiction, *Australasian Journal of Philosophy*, 73, 2, 250–259.

——, 1995c: Imagination as Simulation: Aesthetics Meets Cognitive Science. In Davies and Stone (eds), Mental Simulation. Oxford: Blackwell.

.

——, 1996: Simulation-theory, theory-theory, and the evidence from autism. In Carruthers, P. and Smith, P. (eds), Theories of Theories of Mind. Cambridge University Press.

Currie, G. and Ravenscroft, I., 2002: Recreative Minds: Imagination in Philosophy and Psychology. Oxford: Oxford University Press.

Descartes, R. 1998: Regulae ad directionem ingenii = Rules for the Direction of the Natural Intelligence: A Bilingual Edition of the Cartesian Treatise on Method, (ed./tr.) Heffernan, G. Amsterdam: Editions Rodopi.

Doosje, B., Spears, R. and Koomen, W. 1995: When bad isn't all bad: strategic use of sample information in generalization and stereotyping. *Journal of Personality and Social Psychology*, 69, 642 – 655.

Ericsson, K. and Simon, H. 1993: Protocol Analysis: Verbal Reports as Aata. Cambridge, MA: MIT Press.

Evans, J. and Over, D. 2004: If. Oxford: Oxford University Press.

Fikes, R. and Nilsson, N. 1971: STRIPS: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2, 189 – 208.

Friedman, O. and Leslie, A. 2007: The conceptual underpinnings of pretense: pretending is not 'behaving-as-if'. *Cognition*, 105, 103-124.

Friedman, O., Neary, K., Burnstein, C., and Leslie, A. 2010: Is young children's recognition of pretense metarepresentational or merely behavioral? Evidence from 2- and 3-year olds' understanding of pretend sounds and speech. *Cognition*, 115, 314-319.

Gaut, B. and P. Livingston (eds.), 2003: The Creation of Art. Cambridge: Cambridge University Press.

Gendler, T. 2000: The puzzle of imaginative resistance. *Journal of Philosophy*, 97, 55-81.

Gendler, T. S., and J. Hawthorne (eds.), 2002a: Conceivability and Possibility, New York: Oxford University Press.

Goldman, A. 1992: In defense of simulation theory. *Mind and Language*, 4, 161-185.

——, 2006: Imagination and simulation in audience responses to fiction. In S. Nichols (ed.), *The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction*. Oxford: Oxford University Press.

Goldman, S., Graesser, A., and van den Broek, P. (eds) 1999: Narrative Comprehension, Causality, and Coherence: Essays in Honor of Tom Trabasso. Mahwah, NJ: Lawrence Erlbaum Associates.

Gordon, R. 1986: Folk psychology as simulation. *Mind and Language*, 1, 158-70.

Gordon, R. and Barker, J. 1994: Autism and the 'theory of mind' debate. In G. Graham and G.L. Stephens (eds), Philosophical Psychopathology: A Book of Readings. Bradford Books/MIT Press.

Harris, P. 1991: The work of the imagination. In A. Whiten (ed.), Natural Theories of Mind. Oxford: Blackwell.

——, 1993: Pretending and planning. In S. Baron-Cohen, H. Tager-Flusberg, and D. Cohen (eds), Understanding Other Minds: Perspectives from Autism. Oxford: Oxford University Press.

——, 1994: Understanding pretense. In Lewis and Mitchel (eds), Child's Early Understanding of Mind: Origins and Development. Laurence Erlbaum Associates.

——, 1995: Imagining and pretending. In M. Davies and T. Stone (eds), Mental Simulation. Oxford: Blackwell.

——, 2000: The Work of the Imagination. Oxford: Blackwell Publishers.

Harris, P. and Kavanaugh, R. 1993: Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, 58/1.

Hicks, Robert Drew, 1907. *Aristotle, De Anima*, with translation, introduction and notes,

`Cambridge: Cambridge University Press.

Hjort, M. and Laver, S. (eds) 2004. Emotion and the Arts. Oxford: Oxford University Press.

Hume, D. 1999: An Enquiry concerning Human Understanding. (ed.) Beauchamp, T. Oxford/New York: Oxford University Press.

Ichikawa, J. 2009: Dreaming and imagining. *Mind and Language*, 24, 103-121.

Jarrold, C., Carruthers, P., Smith, P., and Boucher, J. 1994: Pretend play: Is it metarepresentational? *Mind and Language*, 9, 445-468.

Johnson, S., 1751: The Rambler (no. 125).

Johnson-Laird, P. N. and Byrne, R. M. J. 1991: Deduction. Hove: Erlbaum.

——, 2002: Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review*, 109, 646-678.

Kahneman, D. and Tversky, A. 1982: The simulation heuristic. In A. Kahneman, P. Slovic, and A. Tversky (eds), Judgment Under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press.

Kant, I. 1934: Critique of Pure Reason. Macmillan.

Kind, A. 2001: Putting the Image Back in Imagination. *Philosophy and Phenomenological Research*, 62, 85–109.

——, 2011: The Heterogeneity of the Imagination. *Erkenntnis*. 78, 1, 141-159.

Kunda, Z. 1999: Social Cognition. Cambridge, MA: MIT Press.

Langland-Hassan, P., 2011: "Pretense, Imagination, and Belief: the Single Attitude Theory," *Philosophical Studies* 159, 155-179.

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

——, 1994: Pretending and believing: issues in the theory of ToMM. *Cognition*, 50, 211-238.

——, 2002: Pretense and representation revisited. In N. Stein, P. Bauer, and M. Rabinowitz (eds), Representation, Memory, and Development: Essays in Honor of Jean Mandler. London: Lawrence Erlbaum Associates.

Leslie, A. and Thaiss, L. 1992: Domain specificity in conceptual development: neuropsychological evidence from autism, *Cognition*, 43, 225-251.

Leslie, A. and Roth, D. 1993: What autism teaches us about metarepresentation. In S. Baron- Cohen, H. Tager-Flusberg, and D. Cohen (eds), Understanding Other

Minds: Perspectives from Autism. Oxford: Oxford University Press.

Levin, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2013/entries/functionalism/

Lewis, D. 1983: Philosophical Papers, vol. i. New York: Oxford University Press.

——, 1986: Philosophical Papers, vol ii, Oxford: Oxford University Press.

Lillard, A. 1996: Body or mind: children's categorizing of pretense. *Child Development*, 67, 1717-1734.

McCune-Nicolich, L. 1981: Toward symbolic functioning: structure of early use of early pretend games and potential parallels with language. *Child Development*, 52, 785-797.

McGinn, C. 2004: Mindsight: Image, Dream, Meaning. Cambridge, MA: Harvard University Press.

Moran, R. 1994: The expression of feeling in imagination. *Philosophical Review*, 103, 75-106.

Nichols, S. 2004: Imagining and believing: the promise of a single code, *Journal of Aesthetics and Art Criticism*, 62, 129-139.

——,. 2006a: Just the imagination: why imagining doesn't behave like believing. *Mind and Language*, 21, 459-474.

——,. 2006b: Introduction. In S. Nichols (ed.) The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction. Oxford: Oxford University Press.

Nichols, S ., Stich, S., Leslie, A. and Klein, D. 1996: Varieties of off-line simulation. In P. Carruthers and P. Smith (eds), Theories of Theories of Mind. Cambridge: Cambridge University Press.

Nichols, S. and Stich, S. 2000: A cognitive theory of pretence. *Cognition*, 74: 115-47.

——, 2003: Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding other Minds. Oxford: Oxford University Press.

Noordman, L., Vonk, W., and Kempff, H. 1992: Causal inferences during the reading of expository texts. *Journal of Memory & Language,* 31: 573–90.

O'Shaughnessy, B. 2000: Consciousness and the World, Oxford: Clarendon Press.

Perner, J. 1988: Developing semantics for theories of mind: from propositional attitudes to mental representation. In J. Astington, P. Harris, and D. Olson (eds), Developing Theories of Mind. Cambridge: Cambridge University Press.

——, 1991: Understanding the Representational Mind. Cambridge: Bradford Books/MIT

Press.

——, 1993: Rethinking the metarepresentational theory. In S. Baron-Cohen, H. Tager-Flusberg, and D. Cohen (eds), Understanding Other Minds: Perspectives from Autism. Oxford: Oxford University Press.

Premack, D., & Woodruff, G., 1978: "Does the Chimpanzee have a Theory of Mind?" *Behavioral and Brian Sciences*, 1: 515–526.

Radford, C. 1975: How can we be moved by the fate of Anna Karenina? *Proceedings of the Aristotelian Society*, 49, 67–80.

Rips, L. J. and Marcus, S. L. 1977: Suppositions and the analysis of conditional sentences. In M. A. Just and P. A. Carpenter (eds), Cognitive Processes in Comprehension. New York: Wiley.

Rosen, G. 1990: Modal Fictionalism. *Mind*, 99, 395, 327-354.

Ryle, G., 1949: The Concept of Mind. London: Hutchinson.

Sanitoso, R., Kunda, Z., and Fong, G. 1990: Motivated recruitment of autobiographical memories. *Journal of Personality and Social Psychology*, 57, 229-241.

Schroeder, T., and Matheson, C. 2006: Imagination and emotion. In S. Nichols (ed.) The Architecture of the Imagination: New Essays on Pretence, Possibility, and

Fiction. Oxford: Oxford University Press.

Scruton, R., 1974: Art and Imagination, South Bend, IN: St. Augustine's Press.

Sepper, D. L., 1996: Descartes's Imagination, Berkeley and Los Angeles: University of California Press.

Sidelle, A., 2002: "On the Metaphysical Contingency of Laws of Nature," in Conceivability and Possibility, T. Gendler and J. Hawthorne (eds.), Oxford: Clarendon, pp. 309–336.

Soames, S. (Draft, 2014): "Propositions as Cognitive Acts."

Stevenson, L. 2003: Twelve Conceptions of Imagination. *British Journal of Aesthetics,* 43, 3, 238-59.

Van den Broek, P., Lorch, R., Linderholm, T., and Gustafson, M. 2001: The effects of readers' goals on inference generation and memory for texts. *Memory and Cognition*, 29, 1081-1087.

Van Inwagen, P. 1977: Creatures of Fiction. *American Philosophical Quarterly,* 14, 4, 299-308.

Van Leeuwen, N. 2009: The Motivational Role of Belief. *Philosophical Papers,* 38, 2,

219 - 246.

——, 2013: The Meanings of "Imagine" Part I: Constructive Imagination. *Philosophy Compass*, 8, 3, 220-230.

Velleman, D., 2000, "The Aim of Belief," in The Possibility of Practical Reason. Oxford: Oxford University Press.

Walton, K. 1990: Mimesis as Make-Believe: On the Foundations of the Representational Arts. Cambridge: Harvard University Press.

——, 1997: Spelunking, simulation, and slime: on being moved by fiction. In M. Hjort and S. Laver (eds), Emotion and the Arts. Oxford: Oxford University Press.

——, 2006: On the (so-called) puzzle of imaginative resistance. In S. Nichols (ed.) The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction. Oxford: Oxford University Press.

Weatherson, B. 2004: Morality, fiction, and possibility. *Philosophers' Imprint*, 4, 3.

Weinberg, J. and Meskin, A. 2006a: Puzzling over the imagination: philosophical problems, architectural solutions. In S. Nichols (ed.) The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction. Oxford: Oxford University Press.

——, 2006b: Imagine that!. In M. Kiernan (ed.), Contemporary Debates in Aesthetics and

the Philosophy of Art. Blackwell Publishing.

White, A. R. 1990. The Language of Imagination. Cambridge: Blackwell.

Zwaan, R., Magliano, J., Graesser, A. 1995: Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory &Cognition*, 21, 386-397.