IMAGINATION AND EPISTEMOLOGY

by

JONATHAN ICHIKAWA

A dissertation submitted to the

Graduate School–New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Philosophy

Written under the direction of Professor Ernest Sosa

And approved by

____

____

____

____

____

New Brunswick, New Jersey

October, 2008

ABSTRACT OF THE DISSERTATION

Imagination and Epistemology

By JONATHAN ICHIKAWA

Dissertation Director:

Professor Ernest Sosa

Among the tools the epistemologist brings to the table ought to be, I suggest, a firm understanding of the imagination—one that is informed by philosophy of mind, cognitive psychology, and neuroscience. In my dissertation, I highlight several ways in which such an understanding of the imagination can yield insight into traditional questions in epistemology. My dissertation falls into three parts.

In Part I, I argue that dreaming should be understood in imaginative terms, and that this has important implications for questions about dream skepticism. In Part II, I argue that an understanding of the imagination is important for understanding important parts of philosophical methodology—particularly those involving thought experiments. I mean in Part II to be vindicating a great deal of traditional methodology. In Part III, I explore what I take to be a number of deep connections between knowledge and counterfactuals. I defend a form of contextualism in each domain, and argue that inference among imaginings, with its important structural similarities to inference in belief, plays a central role in the epistemology of counterfactuals.

Preface

How can and should epistemologists go about studying questions about knowledge? One way to understand significant movements in the history of epistemology is as providing new approaches to this methodological question. A Cartesian approach asks epistemologists to meditate on *a priori* insights, and to derive conclusions about knowledge from these alone. A Moorean approach offers us, as a starting place, intuitively obvious instances of knowledge, and asks us to work out from these. Later movements recommended further tools to be added to the epistemologist's toolkit: Wittgenstein, Austen, and those who made up the 'Linguistic Turn' directed our attention to everyday use of the word 'knows' and its cognates. Bayesian epistemology suggests that mathematical modeling of probabilities and credences can lead to important insights into the nature and structure of knowledge. More contemporarily, various empirical sciences are being called into the epistemologist's service: psychology (Samuels and Stich 2004), cultural anthropology (Nichols, Stich and Weinberg 2003) and zoology (Kornblith 2003). And, although the stronger claims of the linguistic turn (*i.e.*, philosophical questions are at core questions *about* language) have been mostly abandoned, many epistemologists still think that important insights about knowledge may be achieved via linguistic analysis (*e.g.*, Stanley (forthcoming), Stanley and Williamson 2001).

The central thesis of this dissertation is that a solid understanding of *imagination* should have a place in the epistemologist's toolkit—a grasp of the functioning of human imagination can help us better to understand human knowledge. The case for including it, like that for the value of any tool, comes in the demonstration of its results.

Although philosophical discussion of the imagination stretches at least as far back as

Plato, and finds a place in every major philosophical movement since,[1] imagination has recently come into an unprecedented philosophical focus. This is attributable largely to twentieth-century developments in cognitive psychology, which gave philosophers the resources more precisely to characterize the roles of imagination in philosophically relevant inquiry. The role of imagination in philosophical investigation appears to be extraordinarily widespread: it is generally recognized that imagination has an important role to play in a philosophical understanding of, for instance, the appreciation of fiction, empathy, knowledge of possibility and necessity, hypothetical reasoning, creativity, and mind-reading; according to various views that have been proposed, an invocation of the imagination is also necessary to explain such diverse phenomena as self-deception (Gendler, 2008), delusion (Currie, 2000; Currie & Ravenscroft, 2002; Egan, forthcoming), mental content (Stock, 2003; Weatherson, 2004), belief (McGinn, 2004), the metaphysics of modality (Hill, 2006; Williamson, 2005; Williamson, 2007), and any of the myriad realms in which philosophers have tried to be fictionalists.

As imagination is being asked to play a more and more substantial philosophical role, it is right that the mental phenomenon of imagination should be subjected to increased philosophical scrutiny; it is worthwhile better to understand the activities we talk about and rely upon. In particular, it is important to connect cognitive psychological theorizing about the imagination to the role that imagination is being asked to play in philosophical inquiry.

The project of my dissertation is to elaborate the role that imagination can and should play in epistemology. How can mental processes involving the imagination play justificatory roles in belief formation? Can we come to know propositions by using the

[1] (Brann 1991) provides a thorough historical overview.

imagination? Are there skeptical threats that are particular to questions involving imagination?

At least one instance in which imagination plays an important role in knowledge is already widely recognized: one of the central ways we come to know about the experiences of the people around us is by simulating those experiences in imagination. Alvin Goldman (2006a) has given the most careful and explicit defense of this view. (Goldman focuses on the descriptive claim that we form beliefs about others' mental states by invoking imaginative simulations, backing off from claims about whether we really *know* the contents of the beliefs thus formed. I think a Moorean move is here appropriate: I know that Laura is apprehensive; my belief that she is apprehensive derives from an imaginative process; therefore, such imaginative processes can lead to knowledge.)

Goldman's view about the important role of imagination in mind-reading appears to me to be entirely correct; I have little to add to it, beyond the parenthetical in the previous paragraph. It provides one piece of support, in the form of a case study, for my central claim that epistemologists are well-served to understand the imagination. The project of my dissertation is to provide three more.

Background: Literature Review

I begin with some background explanations of what imagination is, how philosophers have understood it, and what features of it will be particularly relevant for my project.

In this brief section, I outline a few recent philosophical treatments of imagination. In so doing, I hope to clarify the philosophical context with which I am engaging. I make no claims as to the completeness or objectivity of this list—it merely comprises philosophical works about imagination that I have found useful and influential, and highlights what are, for my purposes, the key features of imagination that each posits. I will return to all of them in the later chapters of my dissertation.

Nichols and Stich

In their 'A Cognitive Theory of Pretense,' Shaun Nichols and Stephen Stich (2000) put forward an influential model for understanding imagination in terms of cognitive architecture. Stich and Nichols begin with a 'boxological' framework of cognitive architecture, according to which propositional attitudes consist in syntactic representations playing particular functional roles in the mind. To believe that *p* is to have a mental token representing that *p* in one's 'belief box'; to desire that *p* is to have such a token in the 'desire box'. These boxes are metaphorical; rather than corresponding to particular locations in the brain, they represent classes of functional roles. Representing the mind in this way can help clarify thinking about how various mental states interact; one mechanism will update beliefs in the presence of new evidence, another will take beliefs and desires as inputs and generate actions or decisions, etc.

Nichols and Stich suggest that the best way to understand imagination is to posit, in addition to a belief box and a desire box, an 'imagination box'. (Nichols and Stich originally labeled their box the 'possible worlds' box. For reasons that will emerge in chapter 3, I do not prefer this label, and will follow later work in calling this third box an 'imagination box.') Several features of the imagination box are highlighted: the objects of imagination, like the objects of belief and desire, are propositional representations; imagination differs from belief and desire not in terms of the sort of object it has, but in the functional role of the attitude itself. Nichols and Stich write: 'The functional role of these tokens—their pattern of interaction with other components of the mind—is quite different from the functional role of either beliefs or desires.' (28) This is why I don't, when I imagine that the building is on fire, run outside. In later work, Nichols (2004) emphasizes that although imagination and belief involve distinct functional roles, there are important similarities between them. As he puts it then, 'mechanisms that can take input from the pretense box and from the belief box will treat parallel representations much the same way.' (131) This similarity is meant to explain the roles that imagination can play in affect, and to resolve some *prima facie* puzzling features of typical human engagement with fictions. This similarity forms an important part of the background for my chapter 2, and is a central part of some arguments in my chapters 4 and 6.

A related insight from the Nichols and Stich model is the emphasis on *inference* among imaginings. They (2000) write:

> From the initial [imagined] premise along with her own current perceptions, her background knowledge, her memory of what has already happened I the episode, and no doubt from various other sources as well, the pretender is able to draw inferences about what is going on in the pretense. In Leslie's tea party experiment, for example, the child is asked which cup is empty after the experimenter has pretended to fill up both cups and then turned one upside down. To answer correctly, the child must be able to infer that the cup which was turned upside down is empty, and that the other one isn't, although of course in reality both cups are empty and have been throughout the episode. (119)

And:

> One important part of the story, on our theory, is that the inference mechanism, *the very same one that is used in the formation of real beliefs*, can work on representations in the [Imagination Box] in much the same way that it can work on representations in the Belief Box. (122, emphasis in original)

Imagining that *p* sometimes leads to imagining that *q*, and it does so in a particular, inference-like way. If you imagine that it is raining, this imagining will, in typical cases, lead you to imagine that the ground is wet (or to stop imagining after all that it is raining), in a way not unlike the way you come to believe that the ground is wet when you come to believe that it is raining (unless you go on to reject the belief that it is raining on the basis of a firmer belief about the dry ground). Nichols and Stich posit a cognitive mechanism, ‘The UpDater’, that performs these inferences in both belief and imagination. An appreciation for the parallel nature of these inferences plays a central role in my chapter 4, and represents an important starting point for the project of chapter 6.

Gendler

In some recent work, Tamar Szabó Gendler (2003, 2006b) observes and emphasizes that, although to imagine that *p* is not thereby to believe that *p*, it can, in some cases, confer some of the effects typically associated with the belief. Gendler calls this process ‘contagion’—some associated features of a given imagining have the impacts that a belief with that content would have. For instance, children who are asked to imagine that an empty box has something desirable in it are more likely to look inside the box than are children who are asked to imagine that the box contains something undesirable. Even though they do not believe the contents of the imaginings, those contents have lingering effects on behavior.

We find the same phenomenon in investigation of adults; Gendler points to a study in which normal adult subjects are asked to affix a ‘poison’ label to one of two identical jars of sugar. Even though they are aware of the arbitrariness of the labeling, subjects exhibited a reluctance to eat from the jar marked ‘poison’. Similar themes are present in some of Gendler’s other work. For instance, in her (2000) and (2006a), Gendler suggests that the reason we do not go along with an author’s attempt to make us imagine morally deviant propositions about fictions is because doing so would bring along with it non-imaginative attitudes, such as affective ones, that we find distasteful.

The observation that imagining that *p* can produce the same sorts of affective responses as believing that *p* will figure importantly into chapters 1 and 2.

Goldman

As mentioned above, Alvin Goldman has been a champion of a 'simulationist' approach to mind-reading: according to Goldman (2006a, 2006b), an important part of the story of how we come to understand the people around us is to simulate their experiences in our own imagination. If I want to know what you are experiencing or how you are likely to react, I imagine myself in your position: I simulate, as well as I can, your beliefs, desires, perceptual experiences, etc.

It is tempting to understand this sort of approach as a competitor to the model described by Nichols and Stich: what place is there for *simulation* in their boxological framework? But the central tenets of each approach are consistent, and even complementary. Goldman distinguishes between two types of imagination: he calls them 'supposition-imagination' and 'enactment-imagination':

> To S-imagine that *p* is to entertain the hypothesis that *p*, to posit that *p*, to assume that *p*. Unlike some forms of imagination, S-imagination has no sensory aspect; it is purely conceptual. …
>
> Enactment-imagination is a matter of creating or trying to create in one's own mind a selected mental state, or at least a rough facsimile of such a state, through the faculty of imagination. Prime examples of E-imagination include sensory forms of imagination, where one creates, through imagination, perception-like states. (2006b, 42)

Goldman's S-imagination, I think, fits well with Nichols and Stich's 'imagination box'. We may understand Nichols and Stich as offering a framework for understanding the cognitive mechanisms underlying Goldman's S-imagination. I see no irreconcilable conflicts. Goldman goes on to suggest that S-imagination may be a special case of E-imagination: to S-imagine that *p* is to E-imagine believing that *p*—to simulate the belief that *p*. This may help explain the structural similarities between imagination and belief that Nichols (2004) emphasizes.

(Nichols and Stich (2000) go to some length favorably comparing their own approach to imagination with simulation-emphasizing stories; however, their central argument is that, if suitably developed, simulation-based account would be indistinguishable from their own. (132) So they agree with me that these alternative approaches are consistent. See Goldman (2006a, 46-47) for a related discussion.)

What is particularly valuable, for my purposes, in Goldman's treatment of imagination is the emphasis on empirical understanding of the cognitive mechanisms underlying imagination. Goldman marshals an impressive collection of neuroscientific support for his central claims that imagination importantly involves simulation of other mental states, and that such simulation is an important part of how we understand one another. I have put some of the data that Goldman emphasizes into service in my chapter 1.

Currie and Ravenscroft

Gregory Currie and Ian Ravenscroft's *Recreative Minds* (2002) is a book-length exploration of the nature and philosophical significance of imagination. Like Goldman, they emphasize the simulative aspect of imagination. A 'belief-like imagination' is the simulation of a belief; it's the sort of thing that happens when we suppose that *p* or read that *p* in a fiction. Vision-like imagination is visual imagery, sound-like imagination is auditory imagery, etc. This framework forms an important backdrop for chapters 1 and 2. They argue also that there are states of 'desire-like imagination', not themselves desires, that play an important role in explaining the role of imagination in affect. This thesis has proven controversial; I need not commit one way or the other with respect to it, as I will not invoke desire-like imaginings in this project.

Currie and Ravenscroft, like Nichols and Stich, also emphasize the role of inference in imaginings:

> [T]he reader's inference might be something like this:
>
> Holmes is human. (Something she imagines.)
> All humans are mortal. (Something she believes.)
> So,
> Holmes is mortal. (Something she imagines.)
>
> Here the inference is perfectly conventional; what is notable about it is the commingling of attitudes—belief and imagination—that it involves. And the general principle that governs the making of such inferences is that, where an inference is from premises at least one of which is an imagining, the conclusion will be an imagining as well. (14)

Sosa

In the first chapter of his *A Virtue Epistemology, vol. I*, Ernest Sosa (2007a) argues, first, that contrary to an orthodox view, dreaming does not involve false beliefs, but rather imaginings, and second, that therein lies the solution to dream skepticism: since we do not form false beliefs while dreaming, it is not the case that my present belief could be false and caused by a dream; therefore, dream skepticism poses no epistemic threat. This is, I think, one of the most explicit and interesting treatments of the impact of questions about imagination on epistemology to date. Since Part I of my dissertation is a direct response to this material, I will not further summarize it here. My chapter 1 is a defense of Sosa's claim about the nature of dreaming, and my chapter 2 is an argument against his proposed epistemic impact.

Walton

In his *Mimesis as Make-Believe*, Kendall Walton (1990) develops an explanation of human engagement with art in terms of imagination. Of particular relevance for my project is his engagement with a particular puzzle about affect and fiction: how is it that we have strong emotional responses to fictional characters and events, given that we don't even believe those characters exist, or that those events occur? When Charles attends a horror movie about a slime, he does not believe that there really is a slime, or that he is in danger from one; nevertheless, it seems, Charles is *afraid of the slime*. How can this be?

Walton's own solution is to deny that Charles is experiencing genuine fear; genuine fear, Walton insists, requires the genuine belief that one is in danger. Instead, Walton posits a series of imagination correlates of emotions; we may understand them as analogous to the imagination correlates of beliefs, visual experiences, and the like that have been discussed above. Charles is experiencing *quasi-fear*; a state that is affectively and physiologically similar to fear, but which does not entail any particular beliefs.

Whether this solution strikes one as too radical will probably depend upon his prior commitments about the nature of emotions. Walton is committed to a strong cognitivism about emotions, tying them closely to beliefs. If one is willing to divorce emotions from beliefs to a greater extent than Walton is, then one may say that Charles does experience genuine fear—that belief is not necessary for such emotions. Imagining can play the role of belief in the generation of emotions. This is the line taken by, among others, Currie and Ravenscroft (2002).

I don't mean to take a stand on this question. I highlight the debate because it is directly analogous to one that arises in response to an objection to the view about dreaming I defend in chapter 1.

Summary of the Dissertation

My project is to extend this empirically informed philosophical invocation of imagination to several questions in epistemology. My dissertation comprises three parts, each of which centrally involves an independent epistemic claim. Part I concerns dreaming and dream skepticism. Part II is an imagination-based treatment of thought-experiment methodology. Part III relates counterfactuals and knowledge.

I conclude this preface with a brief summary of each chapter.

Part I: Dreaming

Chapter 1: Dreaming and Imagination

According to a received set of assumptions, dream experience sometimes deceives us. When we dream, we have misleading sensory experiences, and form false beliefs on the basis of them. In chapter one, I argue that this orthodoxy is false. I endorse and expand Ernest Sosa's (2007a) argument that dreams are best understood on an *imagination* model; when I have a dream in which *p*, I do not, while sleeping, believe that *p*; instead, I imagine that *p*. My defense of this claim involves a synthesis of psychological, neurological, and conceptual arguments.

Chapter 2: Imagination and Dream Skepticism

In chapter 2, I consider the epistemic implications of the thesis of chapter 1. According to standard interpretations of dream skepticism, such skeptical arguments invoke the premise that, if I were now dreaming, I would have the same beliefs I actually have. Since the conclusion of chapter 1 is that this orthodox assumption is false, it might be thought that therein lies the answer to, and refutation of, dream skepticism. This is the line taken by Sosa (2007a). I argue to the contrary that the imagination model of dreaming supplies no easy solution to skepticism; that indeed, it suggests a strong epistemic threat, and motivates a novel way to think about skeptical challenges. On this conception, skeptical scenarios need not be scenarios involving false beliefs.

Part II: Thought Experiments

Chapter 3: Thought-Experiment Intuitions and Imagination

Among the ways philosophers sometimes come to know things is by invoking thought experiments. To take a paradigmatic instance, Gettier invited the philosophical community to imagine a fictional scenario about a man whose co-worker owns a Ford; upon doing so, the community came to realize that knowledge is not identical to justified true belief. This methodology has struck some philosophers as suspiciously mysterious: how is it that we can refute a philosophical thesis merely by *imagining* a counterexample? Why should we trust our intuitions in such matters? In this chapter, I develop a treatment of thought-experiment intuitions that assimilates them to judgments of necessity about fictional cases, and argue that skepticism about this ability is unfounded. My account is in much of the same spirit of that of Williamson (2007)—but mine is significantly more conservative with respect to philosophical tradition. In particular, my approach includes a defense the traditional view that in many cases, thought experiment judgments like the Gettier intuition involve *a priori* knowledge of necessity.

Chapter 4: Possibility and Imagination

An important step in most thought-experiment based arguments is a possibility claim: such-and-such situation is *possible*, thus-and-so is true of it, etc. How do we know whether the case we're being asked to imagine is a possible one? A traditional answer invokes imaginability (often under the name 'conceivability'): *p* is possible just in case *p* is imaginable. I take the bald version of this view, strong modal rationalism, to be refuted by scientific essentialism. (This claim is not uncontroversial; I defend it early in the chapter.) However, I do believe that a conservative revision is plausible. I articulate and defend a *weak modal rationalism*, according to which ideal imaginability entails not metaphysical possibility, but conceptual possibility; I also argue that in many cases—including many *a priori* recognizable ones—said conceptual possibility entails metaphysical possibility. I therefore preserve the idea that we may often have *a priori* access to truths of metaphysical necessity. I articulate and defend a novel notion of analyticity in the service of this project.

Chapter 5: Intuitions and Philosophical Methodology

In this brief chapter, I explore some of the implications of the methodology defended in chapters 3 and 4. In particular, I suggest that thought experiment methodology, so understood, need not invoke premises about intuitions; therefore, in a natural sense, it is false that philosophers need to invoke intuitions as a fundamentally important kind of philosophical evidence. In this respect I am in full agreement with Timothy Williamson (2004, 2007). I briefly explore some of the implications of this result for experimentalist critiques of traditional philosophical methodology.

Part III: Knowledge, Counterfactuals, and Imagination

Chapter 6: Knowledge, Counterfactuals, and Imagination

This chapter relates knowledge attributions and counterfactual utterances, and the epistemology of counterfactuals to the performance of knowledge-transmitting inferences in general. The relation's inspiration comes from the observation that a standard way to come to know a counterfactual is to *infer*, in imagination, the consequent from the antecedent. How do you know that *if it were raining, I would have been wet*? You imagine that it was raining, and *infer* on that basis, against the relevant background (i.e., I was outside) that I was wet. This inference is isomorphic to the inference you might perform if you *knew* that it was raining, and came to realize on that basis that I was wet. In my attempt to codify and explain this isomorphism, I defend a particular contextualist semantics for both knowledge attributions and counterfactual utterances. An advantage of my eventual view is that it explains and vindicates widespread beliefs relating knowledge to counterfactuals, such as safety and sensitivity conditions on knowledge. Another is that it resolves skeptical paradoxes, and parallel puzzles about counterfactuals.

Imagination as a Useful Tool.

In my engagement with these three epistemic projects, I hope to establish that the philosophical impact of imagination goes well beyond mind-reading and aesthetics. It can help us better understand central questions in epistemology. I suspect it can help us in other fields, too.

I don't mean here to be arguing that an understanding of imagination is *essential* to answering all of these questions. In particular, I think that it may be possible to restate the central views of Part II without invoking imagination. (This is clearly false for Part I, and strikes me as unlikely for at least parts of Part III.) My claim is that an understanding of imagination is useful in the investigation of these questions, not that it is essential.

Acknowledgments

Unfortunately (not really), the list of people who have helped me to achieve important insights into my dissertation topic is far too long for me adequately to reconstruct. I have been engaging with many of these questions for most of my postgraduate career, and have received invaluable help from literally hundreds of people along the way. Here are a few instances that have stuck particularly in my mind, with apologies to those I have inevitably neglected.

First and foremost, my dissertation committee: Alvin Goldman, Jason Stanley, Brian Weatherson, Tamar Szabó Gendler, and, most of all, my advisor and mentor Ernest Sosa, who introduced me to the significance of imagination for epistemology in his seminar my first semester of graduate school at Brown, and who brought me to Rutgers to write my dissertation. I am extremely privileged to have such an outstanding committee; without their guidance, this work would be considerably weaker.

I am very grateful for my early philosophical development as an undergraduate at Rice University—Eric Margolis, Richard Grandy, and Mark Kulstadt in particular left a great impact on me.

Among the faculty and graduate student colleagues at Brown University, where I began my graduate studies, I am especially grateful to James Dreier, Allan Hazlett, Richard Heck, Chris Hill, Alyssa Ney, Katherine Rubin, Katia Samoilova, Joshua Schechter, Jim Stone, and John Turri. In this group, Benjamin Jarvis requires particular mention—I have worked closely with him on the material in chapters 3 and 4; chapter 3 is a development of a piece we've co-authored (Ichikawa and Jarvis, 2007), and we are currently developing an extension of chapter 4 for eventual publication. Beyond this, conversations with Ben have among the first tests for most of the philosophical theses

I've tried on, and I owe a tremendous amount of the development of both my particular ideas, and my general philosophical outlook, to him.

At Rutgers, beyond my dissertation committee, I am grateful to Saba Bazargan, Nick Beckstead, Heather Demarest, Kate Devitt, Melissa Ebbers, Hilary Greaves, Alex Jackson, Michael Johnson, Peter Klein, Karen Lewis, Barry Loewer, Kelby Mason, Bob Matthews, Zach Miller, Jennifer Nado, Joshue Orozco, Iris Oved, Karen Shanton, Andrew Sepielli, Stephen Stich, Jason Turner, Neil Van Leeuwen, and Dean Zimmerman.

I have had the benefit of discussing material at a number of conferences and events, as well as via email correspondence, over the past several years, and owe a debt of gratitude to many of the philosophers I've interacted with in these environments. In particular, I remember important insights gained from conversations with John Bengson, Michael Devitt, Dorothy Edgington, Andy Egan, Adam Elga, John Hawthorne, Alan Hájek, Peter Godfrey-Smith, Al Mele, Sarah Moss, Shaun Nichols, Jonathan Schaffer, Eric Schwitzgebel, David Sosa, Jonathan Weinberg, Deena Skolnick Weisberg, and Timothy Williamson. I am grateful also to all of them.

Some of the substance of this dissertation has been published independently. Chapter 1 is a slight modification of Ichikawa (forthcoming), forthcoming in *Mind & Language*. Chapter 2 is a development and expansion of Ichikawa (2008), which appeared in *The Philosophical Quarterly*. Chapter 3 is a development of the ideas given in Ichikawa and Jarvis (forthcoming), forthcoming in *Philosophical Studies*.

Table of Contents

Part I: Dreaming

> *Imagine all the people living life in peace. You may say I'm a dreamer, but I'm not the only one.*
>
> —John Lennon

Chapter 1
Dreaming and Imagination[*]

1. The Orthodox View of Dreaming

It is widely assumed, among philosophers, psychologists, and the folk, that dreams sometimes deceive us: that sometimes, we falsely believe that *p* because we have misleading sensory experiences as of *p*, because we are dreaming that *p*. This is why Descartes thinks that the possibility that he is dreaming undermines knowledge of the world around him. 'How often,' Descartes (1986) writes, 'does my evening slumber persuade me of such ordinary things as these: that I am here, clothed in my dressing gown, seated next to the fireplace—when in fact I am lying undressed in bed!' (105) We naturally read 'persuasion' here to involve belief. Contemporary dream psychologist J. Allen Hobson (1999) is explicit:

> What is the difference between my dreams and madness? What is the difference between my dream experience and the waking experience of someone who is psychotic, demented, or just plain crazy? In terms of the nature of the experience, there is none. In my New Orleans dream I hallucinated: I saw and heard things that weren't in my bedroom. I was deluded: I believed that the dream actions were real despite gross internal inconsistencies. I was disoriented: I believed that I was in an old hotel in New Orleans when I was actually in a house in Ogunquit. (5)

I take Hobson's approach to be orthodoxy. The orthodox view is helpfully considered, for my purposes, as involving two views, *percepts* and *beliefs*.

Percepts: Dreaming involves percepts—sensory experiences of the sort we experience during our waking interaction with the world. These percepts are typically misleading; they give us the experience as of perceiving something that is not there.

Beliefs: Dreaming involves beliefs—typically false ones. When we dream that *p*, we believe that *p*. Since in many such cases, *p* is false, dreaming often involves false beliefs.

[*] This chapter is an adaptation of (Ichikawa, forthcoming), forthcoming in Mind & Language.

The role of these two assumptions in standard arguments for dream skepticism should be obvious.

Note that to adopt *percepts* and *beliefs* is not yet to commit to any particular explanation of how it is that percepts and beliefs arise from dreams; Hobson's own view is that they are the product of spontaneous, random activity in the brain stem, but the orthodox picture, as I am specifying it, is also consistent with a more Freudian picture, according to which the percepts and beliefs we experience arise during dreams as the result of higher-level mental activity. The conjunction of *percepts* and *beliefs* do not fully constitute the orthodox view; it may be that on the orthodox view, for instance, the beliefs are *caused* by percepts in the usual ways. But the orthodox view at least entails *percepts* and *beliefs*.

There is room to question the orthodoxy from a philosophical point of view. Kendall Walton (1990) mentions in passing the possibility that dreams are imaginative exercises, and may not involve false beliefs.

> Dreams are beginning to look more and more like games of make-believe, and dream experiences like representational works of art and other props. … Perhaps [the dreamer] doesn't even realize that the propositions in question are *merely* fictional…. Perhaps (as Descartes assumes) dreamers believe what is only fictional in their dreams, as well as imagining it. We needn't decide. (49-50)

Recently, more explicit philosophical challenges have been raised against the orthodox view. Colin McGinn (2004, 74-95) rejects *percepts*; Ernest Sosa (2007, 1-13) rejects *beliefs*. (McGinn (96-112) explicitly endorses *beliefs*; Sosa does not commit one way or the other with regard to *percepts*, but has indicated in conversation that he finds the orthodox picture plausible here.) My project here is to deny both *percepts* and *beliefs*. I will defend a strong version of the *imagination model of dreaming*, according to which dreams typically involve neither misleading percepts nor false beliefs, but instead involve imaginative experiences. My defense will involve a synthesis of philosophical and scientific considerations.

2. Percept and Image

McGinn, Sosa, and I all take it for granted that dreams involve experiences, so I will not attempt to deny *percepts* by denying that there is something it is like to dream. (I therefore reject the approach of Malcolm 1959.) I must instead argue that although dreams do involve experiences, they do not involve percepts—the kinds of sensory experiences we experience when engaging with the world around us. Instead, I will argue that they involve mental *imagery*. Call this claim *imagery*. Visual imagery is the kind of experience one undergoes when imagining what something looks like. Visual imagery is in some ways similar to visual sensation, but it is a different kind of experience. Likewise for auditory, tactile, olfactory, and gustatory imagery. Imagery involves the *simulation* of percept.

2.1. The 'imagery debate' is orthogonal to the imagination model.

There is something called the 'imagery debate' in cognitive science; it is characterized paradigmatically by the many contributions from two of its principal players, Stephen Kosslyn and Zenon Pylyshnyn. Nothing I say should bear on that controversy. The commitments I take on in my discussion of imagery are minimal: there is a kind of mental experience called imagery, usefully thought of as the simulation of percept. It is not identical to percept, but it is in important respects similar to it. At issue in the Kosslyn-Pylyshnyn debate is whether the experience of visual images involves the inspection of a picture-like entity in the head—a question I need take no position on. Even Pylyshnyn, the staunch anti-pictorialist, is happy to admit that the processing of visual imagery involves many of the same cognitive mechanisms as does visual perception, a conclusion Kosslyn has emphasized with compelling neuroscientific data.

Pylyshnyn's main concern seems to be a rejection of a 'Cartesian theater' approach to visual imagery. Indeed, Pylyshnyn (2002) can be read as presupposing the similarity of visual perception and visual imagery when he complains that '[d]espite the widespread questioning of the intuitive picture [Cartesian theater] view in visual perception, this view remains very nearly universal in the study of mental imagery.' (157) In the abstract to the same piece, Pylyshnyn admits that 'it is arguably the case that imagery and vision involve some of the same mechanisms,' while hastening to add that 'this tells us very little about the nature of mental imagery and does not support claims about the pictorial nature of mental images.' In his (1999), Pylyshnyn writes that subjects engaging with imagery 'make the same thing happen in their imagining' that they do when visually scanning a field that is before them. (18)

My minimal conception of imagery carries no commitment to picture-like underlying realizations of images, although it is consistent with such realizations. I assume only that there is a genuine experience of visual imagery, and that it incorporates many of the same processing systems as does visual experience. This is uncontroversial. Alvin Goldman (2006a) provides an effective survey of some of the cognitive scientific bases for this claim. (152-57) For instance, Michael Spivey (2000) has demonstrated that subjects who generate visual images produce saccadic eye movements corresponding to the movements they would make if they were visually examining the real scene corresponding to the imagined one. In a classic study, confirmed by later experiments, Perky (1910) demonstrated that visual perception and visual imagery can interfere with one another. Bisiach and Luzzatti (1978) demonstrated that patients with 'hemispatial neglect'—a tendency to ignore half of one's visual field in perception—ignore, or fail to generate, the corresponding region in their mental imagery. Neuroscientific evidence also confirms that similar processes are at work in vision and visual imagery (see Goldman's

154-55).

This similarity should be unsurprising for conceptual reasons: imaginative states simulate non-imaginative states; the point of visual imagery is to be able to enter into an experience similar to the experience of visual perception.

2.2. Images are not percepts.

Is an image different in kind from a percept? Hume, notoriously, thought not—'impressions' and 'ideas' differ only in degree of *vivacity*. Hume, (1978), I §1 ¶5, writes: 'That idea of red, which we form in the dark, and that impression, which strikes our eyes in sun-shine, differ only in degree, not in nature.' As many authors have observed, this simple picture is surely wrong—vivid images can be more vivid than are faint percepts. But what of the more general suggestion that images differ from percepts only in superficial ways, and not in kind? This is David Sosa's view. Sosa (2006) writes that '[i]mages and percepts are deeply alike; their differences, such as they may be, are inessential to their basic character,' (315) and later characterizes difference between them as 'like the difference between words written in pencil and those written in pen.' (317) Sosa is responding to Colin McGinn (2004), who defends a deeper distinction. McGinn offers a number of characteristic differences between images and percepts; Sosa is arguing that none of the elements in question reflect interesting distinguishing features.

I agree with Sosa that many of the characteristics McGinn lists as typical of images and not percepts are not clear cases of essential qualities of images that are inconsistent with percepthood; nevertheless, I believe that there are sufficient differences to justify treating them as separate categories. At least one element on McGinn's list, I think, captures such differences: the one McGinn calls 'the will'. Imagery, unlike sensation, McGinn suggests, is *subject to the will*. This is not a new idea. Wittgenstein, (1970) whom McGinn cites, writes:

> We do not 'banish' visual impressions, as we do images. And we don't say of the former, either, that we might *not* banish them. (§621, p. 109)

And:

> The concept of imagining is rather like one of doing than receiving. Imagining might be called a creative act. (And is of course so called.) (§633, p. 110)

Sartre (1948) makes essentially the same point in his opus on the imagination:

> A perceptual consciousness appears to itself as being passive. An imaginative consciousness, on the contrary, presents itself to itself as an imaginative consciousness, that is, as a spontaneity which produces and holds on to the object as an image. … It is due to this vague and fugitive quality that the image-consciousness is not at all like a piece of wood floating on the sea, but like a wave among waves. (18-19)

I believe that the idea behind these remarks is correct, and that it does mark an important distinction between percept and image. To imagine is to act—our imagery is in some important sense under our control; this is not so with percepts. David Sosa disagrees—he points out that we do exercise some control over what perceptual experiences we have, since we can choose what to attend to. And imagery, Sosa says, is not always voluntary; sometimes we are frustrated by imagery we'd rather not be experiencing. I may be haunted by a face, or unable to get an annoying tune out of my head.

Sosa is surely right about these possibilities of particular cases, but I do not think that this shows that imagery isn't the product of the will. On the first point, although we do have some control over our perceptual experiences, any such control is indirect; we can take action that we know will result in a changed perceptual experience, but we cannot change our perceptual experience directly. Given the positions we are in, we can no more choose our perceptual experiences than can Sartre's log choose where to be tossed by the river. On Sosa's second point, the fact that sometimes we imagine things we'd rather not be imagining does not show that imagining is not an action and subject to the will.

Unwelcome imagery is more like an unwelcome habit or addiction than an unwelcome set of chains. Wittgenstein's first remark above is particularly apt here—even when our imagery is unwelcome and we cannot banish it, we can *try* to banish it; we know what it is to banish it. We are failing to perform an act. Not so with percepts. The instruction, 'stop having the auditory experience of my voice,' or 'start having the visual experience as of a red square' is a confused one. This is Malcolm Budd's (1989) interpretation of Wittgenstein (105-9); Wittgensteinian or not, I believe this suggestion can distinguish images from percepts. I therefore proceed on the assumption that images and percepts are different in kind.

To establish *imagery*, then, is to establish that the percept-like experiences of dreams are not percepts, but rather images. I have a dream about a kitten. The kitten is gray, and has large green eyes. On *percepts*, the orthodox view, I am having visual color percepts—the same sort of experience I have when I actually see gray kittens with green eyes. According to *imagery*, my sensations are actually of a different sort; they're color *imagery*, the same sort of sensory-like experience I have when I close my eyes and imagine what a gray kitten looks like. (I have experiences corresponding to other sensory modalities, too; if in the dream I am holding the kitten, then I will have the tactile imagery of holding something warm, soft, and slightly shaking.)

3. Considerations Favoring Imagery

The orthodox view is, after all, orthodox. Why adopt *imagery* against tradition? In this section, I highlight a number of considerations in its favor. None provide, I think, deductively valid proofs of *imagery*, but collectively, I take them to suggest a preponderance of evidence in favor of *imagery* over *percept*.

3.1. Dreams don't wake us up.

When I am asleep, a loud noise will typically wake me up. My phone rings, and my sleep is interrupted. The obvious explanation for the causal power of that phone is that it causes auditory experiences—percepts—and those percepts cause me to wake up. Now suppose that I am having a dream in which my phone rings. According to *percepts*, I have the same sort of auditory percept experience that I have when my phone really rings. But if this is the case, we should expect that experience to wake me up. But this is not typically the case; very often, we dream about loud sounds without having our dreams interrupted; usually, actually experiencing loud sounds interrupts our dreams. The proponent of *percepts* owes us an explanation here—she must claim either that it is not the percept but something else that wakes me up when my phone really rings, or she must explain why although such percepts usually wake me, they usually do not when they are the results of dreams. Perhaps she will suggest that it is not my auditory experience that wakes me, but the sound waves hitting my ear. Such alternate explanations do not strike me as particularly plausible, but I recognize this as an empirical issue.

I borrow this argument from McGinn (2004, 80-81).

3.2. Considerations of color favor the imagination model.

Eric Schwitzgebel (2002) has remarked on the philosophical significance of color reports of dreams. Almost all modern Americans report dreaming in color. In the 1940s and '50s, most people believed that visual experience in dreams was a primarily black-and-white phenomenon. This is a difficult discrepancy for *percepts* to accommodate. Did our dreams suddenly become colorized in the late 1950s? What could explain this shift? Schwitzgebel suggests that attributions of dream color track the dominant media in the film industry—people who watch black and white movies describe themselves as dreaming in black and white. It would, I submit, be surprising to learn that the experience of watching movies—something that occupies only a small proportion of even the enthusiastic moviegoer's life—should have such a dramatic and widespread effect on the nature of visual experience in dreams. In the '40s and '50s, dreaming in color was thought to be a sign of insanity! The alternative response on behalf of the orthodox theorist will be to attribute systematic errors to one group or the other. But if we're really so prone to error about fundamental questions about our dream experiences, we begin to lose our grip on why we were tempted by the orthodox theory in the first place.

The imagination theorist need not confront this worry. In visual *imagery*, we can have *indeterminate* color. We can, and often do, call up visual images that have no associated color. This is a difference between images and percepts, and, I suggest, it helps the imagination theorist to explain the color data. Indeed, Schwitzgebel makes something much like this suggestion. (655-56) If our actual experiences in dreams are of indeterminate color, much as the experiences we often experience while imagining fictions are, then it shouldn't perhaps be too surprising if, after the fact, we turn to our common experiences with visual experience in fiction-based, imaginative contexts, to describe our experiences. Those of us who are used to imagining along with color stories, because we see them on television and film, will describe our dreams as colored. So, at least, could go a plausible explanation. So the imagination model may provide the best explanation for disagreement about color sensation in dreams.

These considerations suggest a potentially fruitful line of empirical investigation about the experience of color in dreams—a topic that, as far as I can tell, has received relatively little attention in recent decades.

3.3. Dreaming among children appears to emerge with imaginative capacities.

Studies by dream psychologist David Foulkes, described in his (1999), suggest that, contrary to widespread scientific and folk assumptions, dreaming is a sophisticated cognitive activity that emerges relatively late in human development. Foulkes's study of children suggests to him that children under the age of five dream much less frequently, and much less actively, than do adults. (Just how much less frequently is difficult to determine, since Foulkes finds reason to believe that young children often over-report dreams by confabulating, to please their adult questioners.) Furthermore, dreams develop gradually, beginning with simple, static images and only slowly increasing in complexity. More compelling still is the fact that performance in waking imagery tests is a good predictor of dream development in children, and of dream frequency in adults. The children in his studies who dreamt the least often—and in the least developed ways—were also the ones who performed worst on tests of imaginative ability, even though they had average memory and verbal skills. Foulkes (1999) comes to conclusions that are very friendly to the imagination model:

> From all my data, the suggestion is that dreaming best reflects the development of a specific cognitive competence, indexed by certain kinds of tests of visual-spatial imagination, leading to the conclusion that such imagination must be a critical skill in dream-making. (90)

And:

> To dream, it isn't enough to be able to *see*. You have to be able to *think* in a certain way. Specifically, you have to be able, in your mind's eye, to simulate, at first momentarily and later in more extended episodes, a conscious reality that is not supported by current sense data and that you've never even experienced before. (117)

There is a strong correlation between imagery development and ability and dreaming; *images* might best explain such a correlation. This is admittedly speculative, but I find it suggestive in favor of the imagination model, and worthy of further future study. (This may generalize to provide support for *imagination* as well.)

3.4. Adult case studies confirm connections between dreaming and imaginative ability.

The connections between dream prevalence and imagination appear to continue into adult life. Here, the work of Mark Solms is particularly relevant. Solms's work suggests that the neurological mechanisms necessary for dreaming are not the same ones necessary for sensory experience—rather, they're those needed for imagining. Solms and Turnbull (2002) describe activity in the occipito-temporo-parietal junction as essential for dreaming. 'The occipito-temporal-parietal junction is heavily implicated in the generation of visuospatial imagery,' they write, citing Kosslyn (1994), 'and it is therefore no surprise that it should be implicated in dreaming—which is, after all, a special type of visuospatial imagery.' (203) Solms (1997) surveys historical case studies in which brain trauma resulted in both imagery deficits and cessation of dreaming. (4-19) Summarizing generalizations drawn from case studies, he writes that 'the most robust finding of the present study was the observation that cessation or restriction of visual dream-imagery is invariably associated with a precisely analogous deficit in waking imagery.' (131)

The neurophysiological connections between dreaming and imagination run deeper. Solms and Turnbull (209-11) highlight the apparent structural parallel between dream processing and image processing, with regards to the relationship of each to normal vision.

Visual processing in normal humans occurs in three zones. The first, the primary visual cortex, is connected most directly to the retina, and provides the input into the visual system. Damage to the primary visual cortex results in cortical blindness. However, it does not impair a subject's ability to generate mental imagery; neither does it interfere with dreaming.

The second zone in normal visual processing involves a number of specialized processing tasks, such as color and motion processing and object and face recognition. Damage to this zone results in specific sorts of visual deficiencies, such as impaired color perception or prosopagnosia (the inability to recognize faces). These parallels are also present in visual imagery. (See Kosslyn *et.* A*l.*, 2006, 196-97.)

The third zone, the occipito-temporo-parietal junction, handles higher-level, abstract visual reasoning. Damage to this zone results in 'more abstract disorders that transcend concrete perception: acalclia (inability to calculate), agraphia (inability to write), constructional apraxia (inability to construct), and hemispatial neglect (inability to attend to one side of space).' (Solms and Turnbull, 210) Brain damage to this area eliminates dreaming and also can result in an elimination of visual imagery.

I believe that considerations collectively render *imagery* preferable to *percepts*. At least, it should be considered a plausible alternative to the orthodox view. I will address apparent reasons to think the contrary in §5 below. I begin with the case against *beliefs*.

4. Belief and Imagination

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. The challenge to the orthodox view of dreaming here is parallel to the one above—*beliefs* is rejected in favor of *imaginings*, the view that the belief-like states we take toward the contents of our dreams while dreaming are not beliefs, but rather imaginings.

It is worthwhile to be careful with terminology: *In my dream*, I have many beliefs: *in my dream*, I believe that I am holding the kitten, and I believe that the kitten is purring, and I believe that holding the kitten is the way to attract the attention of my celebrity crush. This alone does not entail *beliefs*; all parties grant that *in the dream* I believe these things; in dispute is whether *in fact* I believe them. The *in the dream* operator should be thought of as analogous to the *in the fiction* operator that is used to explain truths about fictional events. Compare: *in the dream* I am holding a kitten; it does not follow that in fact I am doing so. Depending on how the dream goes, it may or may not be true in the dream that I'm right to think that holding the kitten will make me attractive to my celebrity crush—but whether that is true *in the dream* is independent of whether it is in fact true. We may therefore understand *beliefs* as the claim that my dreaming that *p* entails my believing that *p*. *Imaginings* will have it that dreams are much more like fictions—according to *imaginings*, when I dream that *p*, I do not in general believe that *p*; instead, I imagine that *p*.

Call the belief-like states we experience while dreaming *dream beliefs*; the question before us is whether dream beliefs are beliefs. I will argue that they are not beliefs, but imaginings. Following are several arguments for this thesis.

4.1. The orthodox view faces a dilemma about consistency.

I believe that I am a philosopher, and I've had that belief for some years. Suppose I go to bed tonight and dream that I am an opera singer (who is not also a philosopher). According to *beliefs*, during the half-hour or so in which I am experiencing that dream, I believe that I am not a philosopher. What of my longstanding belief? It seems the orthodox theorist has two choices here. He may admit that I continue to have the longstanding belief—that I am a philosopher—, and temporarily acquire an additional, logically inconsistent belief—that I am not a philosopher. He must admit, then, that during my dream I am experiencing a kind of epistemic irrationality, which I'm only able to detect and resolve upon waking. This seems an undesirable view for at least two reasons: first, dreaming does not seem fundamentally to be an intellectually irrational activity. (I am here in apparent disagreement with (Hobson, 1999, and McGinn, 2004, 113.) After all, agents interested in their own positive epistemic status do not thereby have reason to avoid dreaming, or to take steps, even if there were such possible steps, to dream only truths. The point is not that it is *impossible* to have contradictory beliefs—it's that what we do while dreaming does not seem to fit the model of, for example, instances of self-deception in which subjects might be said to have contradictory beliefs.

Second, the posited resolution upon waking does not seem to match our actual experiences; I do not, upon waking up from my opera dream, introspect and recognize my belief that I am a philosopher, see tension between that belief and the belief that I am an opera singer, and reject the latter belief. When I wake, my dream is over, and my dream beliefs are already finished; I do not have to reject the false beliefs I've acquired. (Note that this is not to deny the familiar experience of uncertainty as to whether an apparent memory is from a dream or real experience. When I introspect and discover an apparent memory that *p*, I may, if the apparent memory is a misleading one, come newly to believe that *p*.)

The other horn of the orthodox theorist's dilemma is to suggest that during my dream, longstanding beliefs that are inconsistent with my new dream beliefs are temporarily abandoned; on this approach, it is not the case that, at 6:00 a.m. this morning, I believed, even tacitly, that I was a philosopher, supposing that I was then dreaming that I wasn't one. I had this belief immediately before bed, and again immediately after waking—these are Moorean facts—but I do not have this belief during the time that I am dreaming.

If the orthodox theorist defends *beliefs* by suggesting that during my dream, I cease to believe that I am a philosopher, that people can't fly, that I am insignificant to Angelina Jolie, etc., then he faces the challenge of explaining this odd fact. We don't typically revise our beliefs drastically and wholesale. Suppose I dream that academic philosophy is, and always has been, a front for an elaborate government conspiracy. Under ordinary circumstances, were I to acquire a belief with that content, it would come gradually in response to mounting evidence; there would be a period where I came to question my long-standing beliefs to the contrary. I'd eventually reject those beliefs in favor of the conspiracy theory. But there is no such transition in dreams. When I dream that philosophers are government conspirators, I do not always, at the beginning of that dream, confront evidence to that effect, leading me to question and overturn my earlier beliefs. I might have a dream like that—a dream in which I gradually discovered the shocking secret of worldwide philosophy departments; but I might just dream this to be the case, without dreaming about myself being confronted by compelling evidence to that effect. Indeed, I might dream that I'd always known about the conspiracy, and was an integral part of it. So if dream beliefs are beliefs, and contradictory longstanding beliefs temporarily disappear, then we have a nightly cases of belief revision that are wildly different from the standard models we encounter in waking life; the orthodox theorist owes us an explanation for such unusual patterns.

On either horn, we see that *beliefs* implies that we have beliefs while dreaming that interact with our longstanding beliefs in very unusual ways, relative to the patterns of waking beliefs.

4.2. Dream beliefs seem relevantly like other dreamt states that do not entail their waking correlates.

Ernest Sosa (2007a) invites us to consider an analogy to intentions. (6-7) Is it the case that when one dreams that she intends that *p*, she thereby intends that *p*? A normative argument, inspired by St. Augustine (1999, book X, ch. 30), suggests not: it is morally reprehensible to intend to do evil, but it is not morally reprehensible merely to dream to intend to do evil; therefore, to dream to intend to do evil is not thereby to intend to do evil. That a person dreams that she intends that *p* may reveal facts about her psychology and her attitude toward *p*, and it may in some cases constitute *evidence* for the claim that she intends that *p*, but it does not itself constitute or entail that intention.

If we agree with Sosa (and Augustine) that dreaming to intend, say, to seduce one's neighbor's wife doesn't imply actually so to intend, then we should find something odd about the idea that the dreamer really believes herself to be seducing her neighbor's wife; the believing and the intending seem to be on a par; to deny that one generates a real truth about the sleeping subject while affirming the other seems *ad hoc*.

If we accept Augustine's normative argument about intention, the burden is on the orthodox theorist to establish why we should understand belief to be importantly different.

4.3. Dream experience is sometimes continuous with imaginative experience.

According to the imagination model, the experiences in dreams are different only in kind from our waking imagistic experiences. The imagination model is supported by the occasional occurrence of experiences that gradually transition from one to the other. Two sorts of cases come to mind. First, consider dreams that develop out of deliberate daydreams. A deliberate daydream is a prototypically imaginative experience—we do not typically believe our waking fantasies to be true. Sometimes, we fall asleep while daydreaming, and the content of the daydream becomes the content of a dream. Insofar as there is a smooth experiential transition between the waking daydream and the sleeping dream, this provides support for the imagination model.

A similar sort of experience sometimes occurs upon waking. If a particularly interesting dream is interrupted, the dreamer might decide to go back to sleep in an attempt to ‘re-enter’ it. When such attempts are successful, there is often a consciously introspectable transitional period in which one experiences the dream with more reflective access than is typical; such experiences can be much like waking imagination.

Admittedly, the considerations raised in this section are anecdotal—I am unaware of any systematic psychological studies of this phenomenon—but perhaps readers who have had similar experiences will find this sort of consideration compelling.

4.4. Conceptual considerations favor the imagination model.

What can a philosopher have to say about the nature of dreams? Beliefs and imaginings are different kinds of representational mental states, and whether dream beliefs are beliefs or imaginings, one might suggest, is an empirical question in cognitive science. According to this suggestion, it would be a mistake from the start to attempt to engage the question of the nature of dreaming from a philosophical approach.

Empirical investigation is certainly relevant to the question of the nature of dreams, and I agree that the question whether dreams involve beliefs or imaginings is in some sense empirical. But I deny the suggestion that a philosopher has nothing of value to contribute here; empirical and conceptual questions are not as distinct as the objection presupposes. The basic neuroscientific facts do not alone settle the question of whether *beliefs* is true; even once the neuroscientific facts about dream beliefs are settled, the question remains whether those states are ones that *count as* beliefs. This is a conceptual question, and one for which philosophical methodology is well suited. My suggestion here is that conceptual considerations also favor *imaginings* over *beliefs*. The way fully to develop this point would be to defend a particular theory of belief, then show that dream beliefs do not meet some of the essential features of belief according to that theory. A defense of a theory of belief is here beyond my scope, but I can mount some pressure against *beliefs* by pointing to ways in which the attitudes in question in dreams are incompatible with some attractive requirements on beliefs.

For instance, some philosophers hold that beliefs necessarily play a certain kind of functional role; that what it is for a mental representation to be a belief, rather than some other propositional attitude, is, at least in part, for it to play a distinctive belief-like role in a subject's cognitive economy. Dream beliefs do not appear to play many of the same functional roles as do prototypical beliefs. They are not connected with perceptual experience in the way that typical beliefs are, and they do not seem to motivate action in the way that typical beliefs do. (McGinn makes much of the observation that dream beliefs engage our affective systems much as beliefs do; I respond to this claim below.) Functionalists are of course free to specify *which* cognitive functions are distinctive of belief, but I am skeptical about there being a satisfactory specification of the functional role of belief that includes dream belief, that does not also include obvious non-beliefs, such as prototypical waking imaginings. What difference between prototypical imaginings and dream beliefs could serve to distinguish between the functional roles of belief and imagination? Beliefhood would have to be consistent with a disconnect from both action and perception, in order to let dream beliefs count as beliefs; but then on what basis would prototypical imaginings be excluded as beliefs?

At any rate, I think there is a clear burden on the orthodox theorist to explain why dream beliefs should count as beliefs, in light of their apparently un-belief-like functional role.

Likewise, if one is tempted by an interpretationalist or dispositionalist theory of belief—I have in mind views like those of Daniel Dennett and Donald Davidson—one will have difficulty granting belief status to dream beliefs. If we observe Laura, who is dreaming that she is in England, on what basis can we ascribe to her the belief that she is in England? She is not behaving as if she is in England—she is not looking to the right before crossing streets. Neither does she give us utterances that are best interpreted as expressions of the belief that she is in England. Indeed, she's exhibiting very little behavior at all, since she is asleep. What is it about Laura in virtue of which she believes she is in England? No answer favoring *beliefs* over *imaginings* seems here to be forthcoming.

I have not argued that there is no plausible theory of belief according to which dream beliefs are beliefs, but I am unaware of one. I believe that the considerations raised here at least put the onus on the orthodox theorist to justify the classification of dream beliefs as beliefs.

Is there a parallel argument against *imaginings*? It is certainly not constitutive of imagining that imaginings be aimed at truth, or that they motivate us in a particular way. What is constitutive of imagination? I suggested in the above discussion of imagery that imagery is in some important sense subject to the will in a way that percepts are not; it is natural to think that imaginings are similarly voluntary. If they are essentially so, this presents a *prima facie* challenge against *imaginings*, since dreams do not seem to be subject to the will in the way that imaginings are. I turn now to objections to the imagination model, starting with this one.

5. Objections to the Imagination Model

5.1. Dreaming, unlike imagination, is involuntary.

Here is an argument against the imagination model: dream experiences—both belief-like and sensory-like—are not typically under our voluntary control, but imaginings and images are always under our voluntary control. Therefore, dream experiences aren't images and imaginings.

The argument is valid and the first premise is indisputable, so the imagination theorist must deny the second premise. Can I do so in light of my §2.2 earlier claim above (p. 6) that a central characteristic of imagination is that it is subject to the will? Yes, because it is possible for something to be subject to the will, and not yet 'under voluntary control'—the annoying song that runs through your head is an example of something like this. It is subject to the will because it makes sense to try to banish it; it is not under your voluntary control because you are unable to succeed. We do not always voluntarily control the things we do; this does not stop them from being things we *do*.

So there is at least conceptual space for the imagination theorist here; he may claim that although dream experiences are not typically under our voluntary control, they are nevertheless subject to the will. But that there is conceptual space for the imagination model is a meager achievement; are there independent reasons to think that dream experiences are subject to the will? I believe there are at least two.

The first reason is that dreams appear to show evidence of design. Random experiences would not be as coherent as dreams are, and dreams experiences can reflect, to some extent, the psychology of the dreamer. (This is no endorsement of any strong Freudianism; that we're more likely to dream about topics that interest us is an instance of this modest claim.) It is natural and reasonable to speak of ourselves as unconsciously *authoring* our dreams; authoring is an active notion. (This observation is also made in McGinn, 2004, pp. 84-85 and Foulkes, 1999, p. 134.)

The second reason in favor of the suggestion that dream experiences are subject to the will is that sometimes, some people have *lucid* dreams—dreams in which the dreamer, aware that she is dreaming, takes active and conscious control over the content of her dream. Percepts and beliefs are never under our active control, so the orthodox model can certainly not describe lucid dreaming; the imagination model fits the experience well. Insofar as lucid dreaming is similar in character to non-lucid dreaming, this provides reason to adopt the imagination model for dreaming in general. In non-lucid dreaming, dreamers create their dreams without recognizing their own agency. In lucid dreaming, dreamers recognize that they are in control of their dream experiences, and are able consciously to direct their dreams. This suggestion is in line with research on lucid dreaming. Stephen LaBerge (2004) writes:

> Strange, marvelous, and even impossible things regularly happen in dreams, but people usually do not realize that the explanation is that they are dreaming. Usually does not mean always and there is a highly significant exception to this generalization. Sometimes, dreamers do correctly realize the explanation for the bizarre happenings they are experiencing, and lucid dreams … are the result. Empowered by the knowledge that the world they are experiencing is a creation of their own imagination, lucid dreamers can consciously influence the outcome of their dreams. (5)

So the relation between imagination and agency poses no objection to the imagination model; indeed, such considerations in conjunction with the imagination model help to explain some interesting features of dreaming.

5.2. The imagination model conflicts with our introspective experience.

Another reason one might reject the imagination model is that it is at odds with our introspective access to dream experience. We are all familiar with dream experience, after all, and dream experiences *seem* to be beliefs and percepts; philosophical arguments cannot overturn the facts to which we have straightforward introspective access. After all, in normal circumstances, it is easy to recognize the difference between belief and imagination, and between percepts and imagery. It is a consequence of the imagination model that many people are mistaken about the nature of dream experience; people often confuse beliefs and imagination, and percepts and imagery, at least in retrospect. So goes the objection.

There are two reasons I do not find this consideration particularly worrying. First, although it's true that many people tend to think of dreams as involving percepts and beliefs, I do not believe that this is because they introspectively reject the imagination model; rather, I think that most people who reflect casually on the nature of dreaming have not considered the imagination model as an alternative to the orthodox theory. As the considerations raised so far have demonstrated, there are subtle distinctions at work between the imagination and orthodox model.

Second, there is precedent for failing to recognize products of the imagination. Psychologists have documented cases in which subjects confuse percepts and imagery. In a classic example from Perky (1910) mentioned in §2.1 above, subjects were asked to visualize objects while, unbeknownst to them, faint projections of those objects were cast in front of them. In many cases, subjects mistook their percepts for images—they perceived real pictures projected onto a screen, and thought that they were imagining them. So it seems that at least under some circumstances, we are prone to mistaking percepts for images. One plausible explanation for this failure in the Perky cases is that subjects' conscious attempts to generate images interfered with their abilities to recognize percepts as such. This fits well with the broader picture we've been working with regarding imagination and agency; perhaps one way we introspectively distinguish images and imaginings from percepts and beliefs is by recognizing our own agency in the generation of the former; subjects in the Perky experiments felt as though they were causing the visual experiences they were having—after all, they were attempting to create such experiences as those experiences occurred—and that's why they mistook those experiences for images. Non-lucid dreamers fail to recognize their own agency in their experiences, which is why they don't always feel like imaginative experiences; lucid dreamers recognize their active roles, and are thereby able consciously to control their experiences. So the imagination theorist has a plausible story to tell as to why dreams do not feel like typical cases of imagination.

This story, incidentally, fits well with Currie & Ravenscroft’s (2002) suggestion that some delusions are *wayward imaginings*, rather than beliefs; subjects fail to recognize imaginings as such because they have a general deficit in recognition of agency. (161)

5.3. Only with beliefs could dreams be as emotionally engaging as they are.

Colin McGinn (2004) accepts *images* but not *imaginings*. He discusses only briefly the suggestion that we ‘no more believe our dreams than we believe our daytime reveries.’ (97) He rejects it on the grounds that without invoking belief, we cannot explain our emotional involvement with dreams:

> The sure test that dreams are suffused with belief is their ability to generate emotions that are conditional on belief, such as fear and elation—with which dreams are full. (112)

The problem for the imagination model, then, can be stated thus:

(1) When I dream that *p*, I experience fear, elation, and other emotions of a certain type.

(2) Emotions like fear and elation, arising from an attitude that *p*, can only arise from a *belief* that *p*. Therefore,

(3) When I dream that *p*, I believe that *p*.

This argument should not be persuasive. This parallels a puzzle in the philosophy of literature involving emotional responses to fictions. Fictions arouse emotions in us without causing belief; we seem to be happy that *p*, even though we do not believe that *p*. This is what philosophers of literature call the ‘Paradox of Fiction,’ and it takes the form of an apparently-inconsistent triad (Radford, 1975, Hjort & Laver, 1997):

(1’) When I read in a fiction that *p*, I experience fear, elation, and other emotions of a certain type.

(2) Emotions like fear and elation, arising from an attitude that *p*, can only

arise from a *belief* that *p*. Therefore,

(3’) When I read in a fiction that *p*, I do not believe that *p*.

It is typically accepted that (3’) is true, so philosophers of fiction generally agree that it is either (1’) or (2) that has to go. We may, with Kendall Walton (1990) and (1997), deny that we really experience fear and elation, but rather experience different, similar states, which he calls *quasi-fear* and *quasi-elation*. Or, we may say with Derek Matravers (1997) and others that belief isn’t *really* necessary for fear; imagination can also play the role that belief often plays in fear, and likewise for the other emotions. It is clear that one of them must be correct.

If we take the latter option to solve our puzzle about fiction, then we have also directly avoided the problem for the imagination model by denying the shared premise (2). If, on the other hand, we insist that these emotions include a cognitive element, denying instead that fictions *really* generate these emotions, then we may very well ask whether dreams *really* generate them either. It will be, perhaps, more plausible if we qualify our denial of emotional involvement in dreams thus: dreams don’t involve emotions, except in the way that fictions do.

It should be noted that McGinn himself draws a tight connection between dreaming and fiction, and he even suggests that dream belief and emotion is in some sense weaker than the waking versions (110-11). But, as emphasized above, he insists that dreams do involve belief, as opposed to some other similar state like imagination.

6. Conclusion

Nothing I have offered here is a conclusive proof of the imagination model over the orthodox model of dreaming. Nevertheless, I trust that I have at least made plausible the possibility of the imagination model as an alternative. Indeed, I think that the considerations raised here push the weight of evidence into the imagination model's favor. Dreams needn't involve false beliefs and misleading sensory experiences. On the imagination model, dreams are very much like vivid daydreams, entered into deliberately and voluntarily. Lose yourself enough in your daydream, and you will feel, in some sense, as if you are really there. That's not to say you falsely believe the contents of the daydream to be true. Our dreams in sleep are, on the imagination model, *like that*.

This conclusion, I think, independently interesting—it may also have broader philosophical implications. As discussed above, dream beliefs provide a valuable test case for theories of belief; those interested in philosophical psychology and the nature of mental content ought to think about the relations between dream beliefs, ordinary beliefs, and the concept BELIEF. The imagination model may also have interesting consequences in epistemology—if dream skepticism is based in the recognition that our beliefs may be false and caused by dreams, the imagination model implies that the key premise for dream skepticism is false. Ernest Sosa (2007a) argues that it is a consequence of the imagination model that *I am not now dreaming* enjoys a similar epistemic status as does Descartes's *I think*; alternatively, one might think that the moral to draw is that false beliefs are less central to skepticism than is widely understood. My exploration of these issues comprises chapter two.

Chapter 2
Imagination and Dream Skepticism[*]

1. The Imagination Model of Dreaming

In chapter 1, I defended the imagination model of dreaming as preferable to orthodoxy. According to the orthodox theory, when dreaming, a subject has misleading sensations, which typically lead to false beliefs. I cited Descartes (1986) as an early proponent of this orthodox view: 'How often,' Descartes wrote, 'does my evening slumber persuade me of such ordinary things as these: that I am here, clothed in my dressing gown, seated next to the fireplace—when in fact I am lying undressed in bed!' (105) By contrast, according to the imagination model of dreaming, when I dream that *p*, I do not have misleading sensory experiences as of *p* and falsely believe that *p*; instead, I have sensory *imagery as if p* and propositionally *imagine that p*. That is, I *simulate* the experience of *p*. On this model, the experience of dreaming becomes more like the experience of fiction, or the experience of a vivid daydream. I don't falsely believe myself to be flying, and to have misleading visual percepts as of the tops of clouds; instead, I'm *imagining* myself to be flying, and calling forth visual *images* of cloud-tops.

The project of this chapter is to explore the epistemic consequences of the imagination model.

As I mentioned in the previous chapter, Colin McGinn (2004) defends the part of the imagination model that describes dream sensory experience as imagistic, but he is careful to deny the parallel suggestion that 'dream beliefs' are not beliefs but imaginings. One reason he offers for denying this latter view is discussed in §5.3 of chapter 1; another concerns us now. McGinn suggests that the imagination model defeats dream skepticism

[*] This chapter is an adaptation of (Ichikawa, 2007), previously published in *The Philosophical Quarterly*.

too easily. '[I]t is precisely the real presence of belief and emotion in the dream,' McGinn writes, 'that gives Descartes's problem bite.' (182) It is my suggestion that McGinn is much too quick here, for at least two reasons. First, it is not clear that, even if the imagination model is inconsistent with dream skepticism, the imagination model should therefore be rejected; perhaps it is skepticism that ought to go. Ernest Sosa rejects dream skepticism on just this basis; I will examine Sosa's argument below. If, as I have argued, our best psychological and philosophical theories established the imagination model as true, it would hardly do to reject those theories on the basis of their conflict with Cartesian intuitions about dream skepticism.

Second, however, it is not clear that the imagination model has the radical anti-skeptical consequence McGinn and Sosa think it does. The central thesis of this chapter is that, far from solving dream skepticism, the imagination model actually brings forward a fairly radical skeptical threat.

2. The Imagination Model and Skepticism

Skeptical scenarios often describe remote possible worlds: you are a brain in a vat, being fed neurological impulses by an evil scientist. It's sometimes thought, in response to these arguments, that there is something illicitly far-fetched about such skeptical scenarios. Ernest Sosa (2007a) puts the point thus:

> Such radical scenarios are often dismissed as ''irrelevant alternatives'' to our familiar common sense. They are alternative, incompatible ways that the world might have been, but not ones that are *relevant*. Why, exactly, *do* they fail the test of relevance? According to one popular view, a possibility is relevant only if it is not *too remote*, only if it might really happen. Possibilities like that of the evil demon or the brain in a vat are said to pose no real threat, being so remote. (2)

The dream scenario seems, among the standard skeptical scenarios, to be uniquely resistant to this move. The dream skeptic does not invite us to worry that the world is drastically different from the way we think it is; he merely points out what we already know: we often dream. The dream scenario is *modally close* in a way that the brain in a vat scenario is not. So dreaming seems uniquely threatening.

Ernest Sosa thinks that the imagination model of dreaming provides an answer to dream skepticism. Here is Sosa's argument. If the imagination model is correct, then we have a quick reply to the skeptic: it is not the case that my belief that I am wearing jeans is threatened by dream skepticism—for if I merely dreamt those jeans, I would only *imagine* them; there would be no false belief.

This does not seem to allay our skeptical worries. But why? My belief that I am in jeans is quite safe from the dream scenario, on the imagination model, since it could not have been caused by a dream—dream considerations provide no nearby possible worlds where I have that belief and it is false. So what worry is left? It seems to have to do with the idea that, even if my belief that I'm wearing jeans couldn't be caused by a dream, a dream could cause an experience that is *subjectively indistinguishable from* the belief that I am wearing jeans. Intuitively, I still can't rule the scenario out—I still can't tell that I'm not dreaming. So now, if the imagination model is correct, instead of worrying that my belief is false, now I have to worry whether my belief is a belief! How am I to know that my internal mental experience is not the result of a dream, given the still-unshaken fact that while I am dreaming, I cannot typically recognize that I'm not awake? Sosa's strategy has two parts. In the negative part, Sosa argues that our inabilities while dreaming pose no in principle threat from indiscriminability to our waking knowledge of being awake. This is supplemented with a positive part, in which Sosa gives a positive argument for the knowability of our wakefulness, given the imagination model. On the negative part, Sosa considers an analogy between dreaming and death. He writes:

> Let us step back. Suppose I could now about as easily be dead, having barely escaped a potentially fatal accident. Obviously, I cannot distinguish my being alive from being dead by believing myself alive when alive, and dead when dead.

> Similarly, I cannot distinguish my being conscious from my being unconscious by attributing to myself consciousness when conscious and unconsciousness when unconscious. But that is no obstacle to my knowing myself alive and conscious when alive and conscious. Might the possibility that we dream not be like that of being dead, or unconscious? Even if one could never tell *that* one suffers such a fate, one can still tell that one does *not* suffer it when one does not. Why not say the same of dreams? (14)

This seems clearly right for the case of death and unconsciousness—I know that I am not dead, even though if I were dead, I wouldn't recognize it. So maybe, Sosa suggests, it's unrealistic to demand that we be able to tell the difference between dreaming and being awake from both sides—maybe one side is enough. Dream states are distinguishable from waking states if we can tell that we are waking, and not dreaming, when we are waking. That we don't have the converse ability while dreaming does not, by itself, undermine our waking knowledge. This much is surely right—subjective distinguishability is not symmetric.

The second part of Sosa's strategy involves a positive story as to how it is that we can know we're awake when we're awake. The idea is that the proposition *I am awake*, like the proposition *I am*, enjoys a special *a priori* status of rational affirmability. As Sosa puts it, each is impossible to affirm falsely. ('We can just as well affirm <I think, therefore I am awake> as <I think, therefore I am>.' (20)) In the case of the affirmation of wakefulness, this is so because, given the imagination model, we do not really affirm things while dreaming. (We are using 'affirm' to mean 'come to believe.' Since, when I dream, I do not form beliefs, I do not, while dreaming, affirm.)

This is a very broad argument against dream skepticism—it does not rely on any contingent facts about how often we dream, or how compelling our dreams feel when we have them; it depends only on the imagination model as a claim about the nature of dreams. If Sosa's argument works, then it will work even in the most epistemologically problematic cases. Here is a scary dream case:

> **Rip the Dreamer**
>
> Rip sleeps, and dreams, twenty hours out of every day. His dreams are very compelling; upon waking, he is invariably surprised to learn that the events of his dream were not actually occurring. Sometimes when he's awake, he reflects on how often things turn out to be just dreams, and he wonders whether his current experiences are dream experiences; he sees no way to tell, other than to wait and see whether he eventually wakes up. His dreams are, for the most part, very realistic: he often dreams about wondering whether he is dreaming; when he does so, he dreams that he sees no way to tell, other to wait and see whether he eventually wakes up.

If the imagination model of dreaming is correct, then during the twenty hours per day when Rip is dreaming, he is not forming beliefs, but rather imagining things. If Sosa's argument works, then it will establish knowledge even for Rip: if Rip, while awake, believes himself to be awake, that belief is safe from the dream scenario; if he were dreaming, he would not have that belief—only an imagining with that content. So if Sosa is right, then even Rip's knowledge is unthreatened by the dream scenario—or at least, it would be if it didn't bring up such doubts in Rip, which sometimes lead him to suspend judgment on propositions he would know, if he'd only stop worry about dreaming and believe them. Rip could conclusively rule out the dream scenario by simply believing it not to obtain. This should strike us as a surprising conclusion about Rip. In what follows, I raise some doubts about the general strategy. I'll suggest that this is not an epistemically permissible move for Rip to make; neither is it for us normal dreamers.

3. Against Sosa's Argument

I grant to Sosa that the fact that some states cannot be recognized introspectively does not imply that we cannot recognize their absence introspectively. The case of death makes this clear. But I maintain that there remain important relevant differences between dreaming and being dead. One difference is that dreaming, unlike death, is an experience; there is something that it is like to dream. So for dreaming, we may sensibly raise the question, *is this experience I'm now having a dream?* Not so for death. Furthermore, the dream states—the experiences of imagery and propositional imagination—are, in an important way, *similar to* the waking states of sensation and belief. There are several reasons to think this. For one thing, as I mentioned above, imaginative states simulate non-imaginative states; the *point* of visual imagery is to play some of the roles of visual sensory experience, in the absence of the relevant external stimuli. The *point* of propositional imagination is to allow us to use a belief-like state without actually believing its contents. Neurological data also supports this similarity claim, as Alvin Goldman (2006a) has emphasized. This undercuts the analogy between dreaming and death. For any experience, we may sensibly ask whether it is dream experience or waking experience, and it is sensible to worry whether we are making mistakes in our answer. One reason this is so is because dream experience is, in some relevant respects, *similar* to waking experience. By contrast, it is not sensible to wonder whether an experience is one in which one exists, or one in which one does not. There is insufficient similarity between existence and non-existence to ground such a worry. This, I suggest, generates at least some pressure against the suggestion that we can know that we're not dreaming merely by citing our beliefs.

What of Sosa's more direct argument, having to do with rational affirmability? Sosa suggests that since I know that I can never falsely affirm that I am awake, and never truly affirm that I am dreaming, it is rational for me to affirm that I am awake; belief in self-wakefulness has the same positive status as does belief in self-existence. Sosa makes a powerful case:

> Consider a *cogito* proposition, such as <I think> or <I am>. Disbelieving is in these cases defective, since self-defeating, for I know that if I take that option I will be wrong. Suspending is also defective, but in a different way. For, I know, about a particular alternative option, that I am epistemically better off if I take that other option, since I will thereby avail myself of a correct answer to my question,

> which I fail to do if I only suspend judgment. Only the believing option is not defective in this sort of way. Only that option is such that I will *not* then be epistemically better off taking either one of the other available options. On the contrary, as I ponder the question whether I think I think and exist, as I epistemically deliberate, the believing option is the only one about which I know ahead of time that my taking it will obviously imply that I am epistemically right in so doing.
>
> On the imagination model of dreaming, <I am awake> shares the noted epistemic status of *cogito* propositions. In its case too, believing is the only epistemically undefective option. Both suspending judgment and believing will share the following feature: that I know ahead of time, as I ponder my question, that I am better off epistemically if I take a particular other option, namely the belief option, since only about that option is it obvious to me now that if I take it I will be right. (18-19)

I am committed to Sosa's central premises: we never wrongly believe ourselves to be awake, and we never truly believe ourselves to be dreaming. And any time I suspend judgment on whether I am dreaming when I was in a position so to believe, I pass up on the opportunity to believe a truth, and thus perform in an epistemically defective way. Furthermore, I agree with Sosa that, now that I know the imagination model to be true, I do know all of this 'ahead of time'. How then can I resist the conclusion that we may rationally affirm wakefulness on no further basis?

Let us back up. 'Affirm,' remember, entails 'believe.' This is why the imagination model precludes falsely affirming that one is not dreaming. We do not come to believe the contents of our dreams when we dream, so we do not affirm our dream events to occur while dreaming. But we do engage in another activity that is in some ways similar to affirmation: we come to *imagine*. Call this activity *quasi-affirmation*. Quasi-affirmation is not affirmation, but it is in many ways similar to affirmation, in just the same ways that imagination is in many ways similar to belief. And from an internal point of view, for the dreamer, quasi-affirmation is importantly like—and indistinguishable from—affirmation. The line, then, is this: one cannot rationally affirm that one is awake, ignoring the possibility that one is quasi-affirming something false.

What are we to say about Sosa's argument? It does not follow, from the fact that I know no affirmation of *p* will be a mistake, that it is rational for me to affirm *p*. If, for all I know, the mental act I'm to engage in will be a false quasi-affirming, then knowledge that I will never affirm falsely is insufficient. Consider an analogy. Manny is deliberating on what to do with the incoming fastball. Suppose he reasons thus: *I have three options with respect to this pitch: I can (a) get a hit, or (b) put the ball into plan and make an out, or (c) take the pitch. Well, option (b) is clearly baseball-defective; it's definitely not a good thing to put the ball into play and get an out. So that just leaves options (a) and (c). And option (a) is clearly preferable to option (c)—after all, (c) will result in either a strike or a ball, but (a) will result in a hit! So I'll get a hit with this pitch.*

Manny is reasoning badly; the proper way to decide what to do with the pitch involves a sensitivity to particular features of the pitch—*e.g.*, whether it is in the strike zone. Manny's transcendental argument ignores such critical factors. Even if it so happens that the pitch is hittable, and Manny succeeds in getting a hit after running this bit of reasoning, the reasoning itself is defective, and the swing is unjustified.

The fallacy, I take it, is transparent: Manny doesn't get to choose whether to make a hit; he only gets to choose whether to take a swing. If he chooses well, his swing will result in a hit. But in order for this to occur, the world must cooperate with him in a certain way. The pitch must be hittable. Although it's true that Manny would never go wrong by getting a hit, it is irresponsible to ignore the possibility that Manny *would* go wrong by *trying* to get a hit and failing. Sometimes, Manny should take the pitch.

I think that Manny's situation is similar to that of the sometimes-dreamer who reasons as Sosa suggests. Getting a hit is like believing yourself awake—it's always successful, when it occurs; making an out is like believing yourself asleep—one never succeeds this way. And taking the pitch is like suspending judgment. When you can't tell whether the world is cooperating with you to a sufficient degree in order to achieve the best state, it's rational to fall back to this safer one.

Admittedly, there are important dissimilarities between Manny getting a hit and Rip believing himself awake. When Manny resolves to get a hit with this pitch, he runs the risk of making a strike or an out—he risks an outcome that is, from the standpoint of baseball, bad. When Rip resolves to affirm himself awake, he does not run the risk of forming a false belief. Perhaps, someone taking Sosa's line will object, the fact that there is no risk of a false belief means that there is no epistemic danger in affirmations of wakefulness.

I do not find this objection compelling. The case of Rip shows that not all epistemic risks are risks of having false beliefs. If Rip is sufficiently reckless, he will, whenever the question comes up, say to himself that he is awake. Sometimes—when he is awake—he will be affirming truly. More often—when he is dreaming—he will be quasi-affirming falsely. Reckless Rip's epistemic behavior is irresponsible; this can only be because false quasi-affirming counts as an epistemic danger.

Here is a similar point. Beliefs in one's own wakefulness are very safe from the dream scenario, for there are no nearby possible worlds where they're false, having been caused by dreams. But safety is only one valuable epistemic state, and the kind of safety that has been established here need not be a great credit to the believer. Consider a variant on the Fake Barn Country thought experiment in which the believer's belief is very safe, if safety is construed along the lines of 'no nearby worlds where I have this belief falsely':

> **Safe Barn in Fake Barn Country**
>
> Henry is driving around in the country, and he sees a bunch of things that look like barns. Henry picks out a barn-looking thing and says to himself, *that's a barn*.

> As it happens, Henry is in Fake Barn Country, where most of the barn-looking things are fake barns. The thing Henry singled out was, in fact, the only real barn in Fake Barn Country, although it had no visible distinguishing marks.
>
> Furthermore, the barn that Henry saw was *modally robust*. In all the nearby possible worlds, that barn was present (and was a barn). The people of Fake Barn Country took that barn very seriously; if it ever fell down, the entire town would have made its restoration their first priority. No one would dream of putting a fake barn *there*.

Henry's belief is true and justified and safe, but it is not knowledge; the differences between this story and the traditional Fake Barn Country story are irrelevant to the evaluation of Henry's belief. This shows that safety—at least when the content of the belief is held constant between neighboring worlds—is not what goes wrong here. Henry is *lucky* to have formed a true belief; he might easily have been looking at a barn façade instead. And Henry isn't so different from my dreamer, Rip, who affirms that he's awake while he happens to be awake. He's lucky, too—he might easily have gone wrong in affirming his wakefulness; he might've been asleep, and quasi-affirmed instead of affirmed. Rip is relevantly like Henry; each belief is deficient for the same reason. This contradicts Sosa's treatment of dream skepticism, on which Rip's waking belief should count as knowledge.

(I have characterized Henry's and Rip's beliefs as safe but lucky; one might alternatively build anti-luck into safety, as in Prichard (2005). This move would undermine the safety of the fake barn case, but it would also undermine the safety of wakefulness: if Rip affirms while waking, he is lucky to have chosen a time when he happened to be awake.)

One might think that there is an important difference between Henry and Rip in that Henry was at risk of forming a false *belief*, while Rip was not. One might admit that both were lucky, but claim that Rip knows and Henry doesn't because Rip's luck didn't save him from a false belief. It's true that this is a difference between the cases, but I do not think it is an important factor in Henry's non-knowledge. Consider Leila:

> **Leila's Demon**
>
> Internally, Leila's mental life is similar to yours or mine. However, Leila is watched over by an unusual demon. Whenever the demon sees that Leila is about to form a belief that is false, he interferes, causing her to imagine the content of that would-be belief instead of believing it. So she often affirms truly, but never falsely; the worst she does is to quasi-affirm falsely. She does this about as often as most people falsely affirm.

Leila never has false beliefs, and her beliefs are extremely safe, but her epistemic position is not thereby flawless. If she knows about her demon, then she may know that for any proposition, she cannot wrongly affirm it; this does not license her rationally to affirm just anything. Rational affirmations are based on sound treatment of available evidence, not on guarantees of non-inaccuracy.

4. Some Remarks About Skepticism

To recap: I have argued thus far that dream experiences are sufficiently similar to waking experiences to justify worries about whether we can know that our present experiences aren't dream ones. This, I suggest, is consistent with the imagination model, according to which we do not believe but rather imagine the contents of our dreams. It follows from this that there is room to worry whether we believe those things that we take ourselves to believe. This threatens to turn a traditional Cartesian epistemology on its head. Descartes wanted to examine all of his beliefs, to see which of them met his high standards; but the imagination model suggests that there is room to doubt not only which beliefs are true, but even which belief-like states are beliefs. Less, then, will end up 'given' than Descartes assumed. This may even threaten our most basic knowledge, in a Cartesian framework.

Here is a completely general version of the skeptical challenge: to have an ideal form of knowledge—*reflective* knowledge—one must be able to defend one's belief against salient alternative explanations. In order reflectively to know, one must know (or be in a position to know) that he knows. To know that one knows, one must know that one believes, since knowledge transparently entails belief. But to know (again, in this reflective way) that one believes requires ruling out all relevant alternatives—including, in this case, that one does not believe, but merely imagines. An imagining is not the sort of thing that can qualify as knowledge. So, here is the general argument against reflective knowledge of any *p*. I am assuming, with philosophical tradition, that we have no internally accessible test for wakefulness; I will question this assumption in my conclusion.

(1) To have reflective knowledge that *p*, I must be in a position reflectively to know that I know that *p*. (Assumed)

(2) To be in a position to reflectively know that I know that *p*, I must be in a position reflectively to know that I *believe* that *p*. (Assumed)

(3) In general, if there are nearby possible worlds I cannot rule out where not-*q*, then I cannot know that *q*. (Assumed)

(4) If there are nearby possible worlds I cannot rule out where I do not believe that *p*, then I am not in a position reflectively to know that I believe that *p*. (Instance of 3)

(5) There are nearby possible worlds I cannot rule out where I do not believe that *p*. (Implication of imagination model)

(6) I am not in a position reflectively to know that I believe that *p*. (4, 5, *modus ponens*)

(7) I cannot have reflective knowledge that *p*. (2, 6, *modus ponens*)

If I am right about this, the imagination model of dreaming, along with the assumption that we cannot distinguish dreams from wakeful life on substantive grounds, is devastating for any Cartesian-style anti-skeptical project. Reflective knowledge is impossible, because we have no reliable way of recognizing our beliefs. Any attempt to begin an epistemology by examining our beliefs is undermined by dream skepticism.

What is new about this argument, other than its novel structure, is its wide scope. This argument provides a perfectly general argument against knowledge of any proposition—even those propositions that are generally thought to have a special, *a priori* status. The *cogito*, for instance, does not appear to have a special immunity from this argument, the way it does from the more traditional Cartesian worries.

This is not, of course, to embrace skepticism; it is a familiar point that Descartes seems to have set the bar for knowledge much too high—it is a commonplace that our ordinary empirical beliefs do not reach it. I suggest that, on the imagination model, even the *cogito* may not reach it. However, there may be an anti-skeptical complication. If, as is plausible, we retain *tacit* beliefs while dreaming, it is not the case that dreaming provides nearby worlds where we fail to believe in our existence—so premise (4) may not be true for propositions like the *cogito* after all. A careful treatment of this point would involve a theory of tacit belief, a story about what sorts of things we tacitly believe, and an exploration of the principles relating tacit belief to reflective knowledge.

I think that these considerations invite an alternative way to think about skepticism. On a traditional understanding of skepticism, the relevant threat is that we have false beliefs. I think that considerations raised here demonstrate that this is an incomplete characterization. For I agree with Sosa that since the imagination model is correct, there is no threat from dreaming of our having false perceptual beliefs—that the dream possibilities we're to worry about aren't possibilities in which we believe falsely. And yet I maintain, given our inability to distinguish dream experience from waking experience, things have gotten worse, not better, for the anti-skeptic.

Here is a suggestion: the real specter of skepticism isn't that we have false beliefs. The real specter of skepticism is that *these things I believe*—here I am referring *de re*—are false. That is, that I am not really a philosopher; that Neil Armstrong did not really walk on the moon; that there really is no external world. A possible world in which there are no tables is, if I cannot rule it out, an epistemic danger; adding to the scenario that I do not *believe* there are any tables is, if I still cannot rule it out, no less dangerous—such a stipulation simply adds that another of my actual beliefs is false! This is the lesson of Rip and Leila; it is not enough to dispel a skeptical worry to observe that, if the worry obtained, one would not have a false belief.

Thinking of skepticism in this way may also help explain what is wrong with Hilary Putnam's (1981, 1-21) argument against brain-in-a-vat skepticism via semantic externalism: perhaps, if I were a brain in a vat, most of my beliefs would be true. Nevertheless, if I were a brain in a vat, I would not have hands; therefore, the brain-in-a-vat scenario does still, if I cannot rule it out, threaten my apparent knowledge that I have hands.

5. Conclusion

I am not a skeptic; I think that I know that I have hands, that Neil Armstrong walked on the moon, and that I am not now dreaming. But I think that there is a serious skeptical challenge raised by dreaming—and particularly by the imagination model of dreaming. My project here has been to articulate that challenge in as strong a form as possible. The correct response to that challenge, I think, will be based in a solid understanding of the role of imagination in dreaming, the nature of imagination, and the epistemology of imagination: of crucial importance is how we can and do know the difference between our beliefs and our imaginings. There may be some reason for optimism in the observation that we do not, in typical (waking) situations, confuse the two; we do, for the most part, manage to know which things we believe and which ones we imagine. An important epistemic project is to investigate how we are able to achieve this discrimination; we may hope that what explains this knowledge can also explain our knowledge of our own wakefulness. I hope to explore this project in future work.

It should be clear that the imagination model of dreaming importantly changes the shape of debates about dream skepticism. Once we deny the Cartesian assumption that we believe what we dream, what epistemic consequences follow? I have argued against Sosa's suggestion that affirmations of wakefulness are easily justified by a transcendental argument, and instead argued that, absent a basis for distinguishing waking life from dream, the imagination model poses an important epistemic threat: not the threat that we might have false beliefs, but the threat that *these* propositions—here I am pointing to the ones we believe—may be false (and, maybe, not after all believed). I closed with a brief suggestion as to how we ought to proceed in pursuit an epistemology of wakefulness. Whether or not this tentative suggestion turns out to be right, it should be clear that there is an important place for psychological questions about the imagination in future investigation of dream skepticism.

Part II: Thought Experiments

> *Imagination will often carry us to worlds that never were. But without it we go nowhere.*
>
> —Carl Sagan

Chapter 3
Thought-Experiments Intuitions and Imagination[*]

1. Introduction

Among the ways we come to learn things is a method involving thought experiments. I imagine a state of affairs, and I make an intuitive judgment about that state of affairs, and come to learn something new on that basis. Perhaps I confirm or falsify some general theory. Here is a short fiction illustrating the familiar point:

> Professor McStory was teaching an epistemology course. 'It's surprisingly difficult,' he said, 'to provide an analysis of knowledge.' A student in the back row raised his hand. 'Yes, Brian?'
>
> 'I don't see why we should think it's so hard to provide an analysis of knowledge,' Brian said. 'I think I know what knowledge is—it's justified true belief.' Professor McStory marveled privately at how conveniently the dialectic was progressing. 'Let me tell you a story.' The class leaned forward attentively. 'Listen to this, and see if you think that knowledge is justified true belief.
>
>> Joe had left his watch at home, and he wanted to know what time it was. He saw a clock on the wall. As a proficient reader of clocks, Joe had no difficulty in determining that the clock read 10:15. "Good," thought Joe, "I still have fifteen minutes until Mr. Pumbleton will be expecting me." Joe had arranged an important meeting at 10:30.
>>
>> However, things were not as they appeared, for Joe had been looking at an inaccurate clock—it was 15 minutes slow. It was already 10:30. But fate smiled on Joe that day—for due to a careless error on the part of Mr. Pumbleton's secretary, Mr. Pumbleton thought Joe's appointment was at 10:45. So Joe's belief about how much time he had until Mr. Pumbleton would consider him late was true after all.
>
> Think about Joe's belief about how much time he had,' Professor McStory suggested. 'It was justified, and it was true, but intuitively, it was not knowledge. So knowledge isn't justified true belief.' Brian and the rest of the class thought about the story and conceded that Professor McStory was right. Justified true belief wasn't the same as knowledge. The end.

[*] This chapter is based on the material in (Ichikawa and Jarvis, forthcoming), forthcoming in *Philosophical Studies*.

Professor McStory's presentation is a fairly typical example of the role of thought experiments in standard philosophical methodology. Thought experiments are invoked in non-philosophical areas, too. In the natural sciences, Galileo, Einstein, and Schrödinger made critical use of well-known scientific thought experiments. (See Gendler 1998.) But it's sometimes thought that philosophical thought experiments are importantly different from scientific (or economic, or poker-based, etc.) thought experiments; *philosophical* thought experiments, it is said, crucially rely on the invocation of *intuitions*.

But there is apparent reason to be skeptical about such invocation of intuitions; one may worry that to posit a faculty that can accurately judge non-actual thought-experiment situations is unduly mysterious. If we think that we can learn about knowledge (or morality, or reference, or personal identity, etc.) by invoking intuitions, then we must think that there is some reliable connection between our intuitions and the target phenomena. But, skeptics may press, no explanation for such alleged reliability, or reason to have such faith in ourselves, is forthcoming.

One possible response to this pressure is to change the subject—to suggest that McStory isn't teaching his class about knowledge itself—instead, he's teaching students about the *ordinary concept* KNOWLEDGE, or the meaning of the English word 'knowledge', or some other mind-dependent target of inquiry. If these were the targets of philosophical inquiry, then the role of intuitions in revealing their features might be rather straightforward. Much twentieth-century analytic philosophy followed this path down the 'linguistic turn'; some recent work from contemporary philosophers is also squarely in this tradition—Alvin Goldman (2007) is explicit in taking the subject matter of philosophy to be ordinary concepts, at least partially on the grounds that doing so is a way to avoid skeptical worries about intuition.

It is clear that there are questions worth asking about the concepts like our commonplace concept KNOWLEDGE, and it is not implausible that some of those questions are of philosophical interest. But it is a serious misconstrual of philosophy to claim with generality that philosophical questions are questions about concepts (or words, etc.). In particular, Professor McStory's aim is not to teach his students anything about the English word 'language', or about their own concepts, but about knowledge itself. Many philosophical questions are about objective phenomena; standard methodology permits thought experiments even in these investigations. (For similar statements, see each first chapter of Kornblith, 2003 and Williamson, 2007.)

How could we be justified in trusting *intuitions* to tell us about objective features of the world? From a naturalistic point of view, to suppose that we have such abilities can look like vain superstition. There is now an extensive literature challenging the role of intuitions in philosophical inquiry. Naturalistic skeptical arguments threaten much of traditional philosophy. If the critics are right, then Professor McStory is presupposing a problematic philosophical methodology. Far from helping them see into the nature of knowledge, he is confusing and indoctrinating his students with philosophical dogma. We cannot trust our intuitions to tell us whether something is a case of knowledge or not; we must do empirical investigation of some sort.

More recently, Timothy Williamson (2004, 2005, 2007) has offered a sort of a defense of Professor McStory's strategy. On Williamson's view, there is no special faculty of intuition that we must invoke in order to judge that Joe doesn't know. Instead, the 'intuition' we form is just a straightforward judgment, no different in kind from the naturalistically innocuous judgments we make every day. Although Williamson is aptly read as defending McStory's invocation of a thought experiment, and I agree with much of Williamson's broader picture, I do not think that it is right to think of Williamson's particular account of thought experiment intuitions as vindicating anything much like a traditional understanding of philosophical methodology. The judgments with which Williamson identifies ‘intuitions’ are scarcely recognizable as the things traditional philosophers had in mind. In particular, on Williamson’s view, the propositional contents of the intuitions in question are contingent. This is problematic from a traditional standpoint for two reasons. First, tradition has it that intuitions like the Gettier intuition have necessarily true contents; Williamson’s counterfactuals, however, will vary in truth value between possible worlds. Second, the standard view has it that intuitions like the Gettier intuition can be known *a priori*; on Williamson’s account, this will be impossible. Indeed, I will argue that on Williamson’s account, not only will intuitions like the Gettier intuition not be knowable *a priori*, they will often not be knowable at all. Indeed, many instances of Gettier intuitions will be false, on Williamson’s view.

The skeptics charge that traditional philosophical methodology should be abandoned as superstition; Williamson’s formulation makes some progress in ‘de-mystifying’ the intuitive judgments, but leaves them as problematic *a posteriori* counterfactuals. My project will be to find a middle ground for the traditional view of thought experiments and intuitions. I think that we can give a naturalistically innocuous account of thought-experiment intuitions, just as Williamson hoped; furthermore, I will argue that Williamson is wrong in thinking that we must give up claims of necessity and *a priority* to do so. I seek an account of thought-experiment intuitions that vindicates traditional philosophical methodology, and *also* preserves these traditional features. A key to my solution, as it is to Williamson’s, will be that the application of concepts to imagined events will be explained by just the same cognitive faculties that explain the application of concepts to perceived events. Any non-skeptic about the latter (and we should all be non-skeptics about the latter!) should also be a non-skeptic about thought-experiment intuitions.

I begin with a restatement of Williamson's argument, then attempt to establish a suitable alternative. In the final sections of the chapter, I argue that this formulation, like Williamson's, provides a plausible explanation for the relevant reliability. I will present Williamson's approach in §2, and criticize it in §3. §4 is a clarification of my project. My own view is given in §§5-6, and defended from a kind of objection in §7. The final sections, §§8-12, are devoted to a defense of the each step in the standard methodology as both warranting and *a priori*.

The Gettier intuition is useful as a paradigm, and I will follow recent tradition and focus on it. Much of what I say should generalize to other invocations of thought experiments in traditional philosophical methodology. (See §10 below for discussion of such generalization.)

2. A Williamsonian Argument

How shall we formalize the Gettier argument? Here is a first stab:

$K(x, p)$: x knows that *p*.
$JTB(x, p)$: x has a justified true belief that *p*.
$GC(x, p)$: x stands to *p* as given in the (interpreted) text of the Gettier story[1]

(1) $\Diamond \exists x \exists p\, GC(x, p)$

(2_n) $\underline{\Box \forall x \forall p\, [GC(x, p) \supset (JTB(x, p)\ \&\ {\sim}K(x, p))]}$

(3) $\Diamond \exists x \exists p\, (JTB(x, p)\ \&\ {\sim}K(x, p))$

In English:

(1) It's possible for some x to stand to some *p* as given in the text of the Gettier story.

(2_n) Necessarily, if x stands to *p* as given in the text of the Gettier story, then x has a justified true belief in *p* that isn't knowledge.

[1] Williamson puts it thus: *x stands to p 'in the relation described by the Gettier story*' (2007, 184). I am characterizing Williamson's GC relation in terms of *texts* of Gettier stories, which is clearly what Williamson has in mind, because there is an important difference between stories and texts. I will elaborate this difference below.

Therefore,

(3) it's possible for someone to have a justified true belief that isn't knowledge.

In this formulation, (2_n) represents the Gettier intuition. And this Gettier intuition is a necessity claim, just as traditional philosophers think it is.

But Williamson argues that this cannot be the correct formulation of the Gettier intuition. For (2_n), says he, is false. The key to seeing this is that the texts of Gettier stories are inevitably underspecified. It is possible, consistent with the text of the story, to insert pieces that will prevent the story from describing a case of justified true belief that isn't knowledge. We can describe cases in which the given Gettier text is true, but there is no case of non-knowledge justified true belief (hereafter NKJTB), as *bad* Gettier cases. Gettier cases with NKJTB are *good* ones.

Following are two ways we could understand Professor McStory's story from the introduction as a bad Gettier case by adding consistent bits to it:

(a) Joe had left his watch at home, and he wanted to know what time it was. He saw a clock on the wall. As a proficient reader of clocks, Joe had no difficulty in determining that the clock read 10:15. **Joe ignored the dozens of other clocks decorating the wall, each of which read '10:30'.** 'Good,' thought Joe, 'I still have fifteen minutes until Mr. Pumbleton will be expecting me.'

(b) As a proficient reader of clocks, Joe had no difficulty in determining that the clock read 10:15. 'Good,' thought Joe, 'I still have fifteen minutes until Mr. Pumbleton will be expecting me.' **Joe knew that Mr. Pumbleton was watching him via closed-circuit television and would accurately predict that it would take him fifteen minutes to reach his office from his current location.**

Both (a) and (b) describe bad Gettier cases—in each case, the original literal text is satisfied, but Joe does not have NKJTB. In (a), Joe's belief fails to be justified; in (b) his belief qualifies as knowledge. Williamson points out that, given the possibility of bad Gettier cases, (2_n) cannot be right. Sometimes, when x stands to *p* as in the Gettier text, there is no NKJTB. Williamson is surely correct about this point; (2_n) is false, and this renders the first stab formulation unacceptable.

A philosopher's natural move here is to tweak the story. We'll come up with a new Gettier text, and do so with extra care to 'close the loops.' If one is clever and careful enough, one might be able to generate texts that are not susceptible to bad satisfaction. Even if this is true, however, it misses the force of Williamson's observation. The problem is not that it is impossible to construct 'invincible' Gettier texts—it's that most of the texts we actually use, *and appropriately use*, are not invincible. Professor McStory might have been able to construct an airtight story, but he didn't have to. The text given at the start of the chapter works perfectly well to generate the Gettier intuition, but the first stab formulation of that intuition, (2_n), is false, because the text can be satisfied in bad ways.

So (2_n) must be weakened. Here is Williamson's version of the argument:

(1) $\Diamond\exists \mathrm{x}\exists p\ \mathrm{GC}(\mathrm{x}, p)$

(2_{cf}) $\underline{\exists \mathrm{x}\exists p\ \mathrm{GC}(\mathrm{x}, p) \mathrel{\Box\!\!\rightarrow} \forall \mathrm{x}\forall p\ [\mathrm{GC}(\mathrm{x}, p) \supset (\mathrm{JTB}(\mathrm{x}, p)\ \&\ {\sim}\mathrm{K}(\mathrm{x}, p))]}$

(3) $\Diamond\exists \mathrm{x}\exists p\ (\mathrm{JTB}(\mathrm{x}, p)\ \&\ {\sim}\mathrm{K}(\mathrm{x}, p))$

In English:

(1) It's possible for some x to stand to some *p* as given in the text of the Gettier story.

(2_{cf}) If some x were to stand to some *p* in a way satisfying the (literal interpreted) text of the Gettier story, then anyone who satisfied the text of the story with a proposition would have NKJTB.[2]

(3) So, it's possible to have NKJTB.

[2] The counterfactual Williamson is really after is the one expressed by the English, 'if someone were to stand to some proposition in a way satisfying the text of the Gettier story, then *he* would have NKJTB of *it*.' He ends up with the more awkward (2_{cf}) because it is difficult to formalize counterfactuals with this pattern of pronoun anaphora (so-called 'donkey counterfactuals'). See Williamson (2007, 195-99).

This formulation of the Gettier argument, like the first, is valid.[3] It also avoids Williamson's objection to the first argument, because the mere possibility of bad cases does not falsify the counterfactual. It's *possible* to stand to a proposition in a way matching the text's description of Joe's relation to *I have fifteen minutes* without NKJTB—but, Williamson says, the relevant counterfactual does not pick out the bad cases. Rather, he suggests, *if* someone *were* to stand to a proposition in the way matching the text, it *wouldn't* be bad, even though it's possible to be in such a position in a bad way. *If* someone *were* to stand to a proposition in a way matching the story, it would be a case of NKJTB.

3. Worries for Williamson's Interpretation

There are at least two reasons to be uneasy with Williamson's treatment. The first is one I've mentioned already: Williamson's interpretation of the Gettier argument renders the intuition as a judgment of an *a posteriori* counterfactual; (2_{cf}), which is meant to correspond to the Gettier intuition, can be at best contingently true. This is so because the truth of a counterfactual depends on, in Lewis and Stalnaker's formulations, characteristics of the *nearest* possible worlds where the antecedent is true.[4] And which worlds are the nearest worlds depends on what the actual world happens to be like. The counterfactual can be known at best *a posteriori*, because knowing what the actual world happens to be like requires empirical investigation. So warrant for the Gettier intuition, on Williamson's view, will involve a great deal of empirical knowledge about the actual world. If the Gettier intuition is like this, then it is not the sort of thing that traditional

[3] Or so, with Williamson, I am willing to assume. Williamson's formulation is invalid if there are substantive counterpossible conditionals that do not automatically make impossible worlds more distant than possible ones. Thanks to Brian Weatherson for this observation.

[4] I defend a non-Lewisean account of counterfactuals in chapter 6. But it, like Lewis's, is one in which counterfactuals with contingent antecedents and contingent consequents are often contingent and *a posteriori*.

philosophy takes it to be. Williamson's account leaves intuitions as mere judgments about contingent matters of fact. As such they cannot be knowledge *a priori*.

In the context of Williamson's greater project, my alleged worry may be seen as a virtue: by denying that the Gettier intuition is known *a priori*, we assimilate philosophical judgments to judgments in general. So the contingent, and *a posteriori*, nature of the content of the Gettier intuition will be of no concern to those generally skeptical about apriority in philosophy—only to someone who, like me, is interested in vindicating that particular traditional feature of thought-experiment intuitions.

But there is a more direct difficulty with Williamson's formulation: it is not only that (2_{cf}) not be known *a priori*; for many—perhaps most—Gettier thought experiments, it cannot be known at all. Indeed, in some cases, it is probably false. Williamson's version of the Gettier intuition is, *if some person were to stand to some proposition in a way satisfying the text of the Gettier story, then anyone who satisfied the text of the story with a proposition would have NKJTB.* This counterfactual is false when the nearest worlds where someone satisfies the text include someone doing so in a bad way. Suppose that, perhaps unbeknownst to the Professor, there is someone in the actual world who satisfies the text of Professor McStory's story—and that he does so in one of the bad ways specified above. Then, in the nearest world where someone satisfies the text—the actual world—it is not the case that everyone who satisfies the text has NKJTB. Williamson's (2_{cf}) is false.[5]

This cannot be right. The appropriateness of McStory's thought experiment does not depend on there being no one in the actual world who satisfies the text in a bad way. On

[5] This is not a mere feature of the particular awkward attempt to formalize the English counterfactual (see fn. 2); it is also false, in the case described, that 'if someone were to stand to some proposition in a way satisfying the text of the Gettier proposition, he would have NKJTB.'

Williamson's formulation, if there is such a person, McStory's students are misled, and *falsely* believe (2_{cf}). If they go on properly to reason to the conclusion that knowledge is not justified true belief, then they have come to a true belief about the nature of knowledge by competently deducing it from a false (perhaps justified) belief. And as we well know, competently deducing true beliefs from false justified beliefs is a standard formula for generating NKJTB; so Williamson's view implies that in this case, McStory's students do not know that knowledge is not justified true belief. A defective thought experiment has generated a Gettier case about the analysis of knowledge! This is unacceptable.

A natural move is to weaken Williamson's counterfactual; we replace the universal quantifier in the consequent with an existential:

$$(2_{cfw})\quad \exists x \exists p\, GC(x, p) \;\Box\!\!\rightarrow \exists x \exists p\, [GC(x, p) \supset (JTB(x, p)\ \&\ {\sim}K(x, p))]$$

This does not avoid the objection, for this counterfactual is false in worlds where exactly one person satisfies the story, and does so in a bad way. Furthermore, an additional objection arises as this weakening doesn't allow for straightforward disagreement of intuitions. Someone who has the Gettier intuition, (2_{cfw}), won't disagree with some (nonstandard) person who has the opposing intuition about the case and hence accepts $(*) \exists x \exists p\, GC(x, p) \;\Box\!\!\rightarrow \exists x \exists p\, [GC(x, p) \supset (JTB(x, p)\ \&\ K(x, p))]$, as ($2_{cfw}$) and (*) are consistent.

4. 'The' Gettier Reasoning

It may be helpful at this stage to clarify a methodological point. What am I doing when I try to explain *the* Gettier argument? Ernest Sosa has raised, in conversation, an objection to the two-premise structures that Williamson uses, and that my eventual view will also use. Sosa suggested that such structures don't do justice to the simplicity of the Gettier intuition. Really, in having the Gettier intuition, one just comes to see that one way to satisfy the text of the Gettier story involves someone having justified true belief without knowledge. Sosa's suggestion may amount to the proposal that people directly apprehend (3). Moreover, while (3) is no doubt necessarily true, necessity doesn't enter into its content. The propositional content of the Gettier intuition, then, is much simpler in form than any of the considered versions of (2).

I am not confident that Sosa's suggestion represents a genuine alternative to the two-premise structure. After all, we will need some account of how the thought experiment generated (3). Presumably, it did so through guiding one to consider a particular sort of possibility. But it's not enough merely to have in mind this determinate possibility; one has to see that this possibility is one in which there is an instance of justified true belief without knowledge. Of course, this process is very close to the kind of structures that Williamson is engaging with, and it will be very closely related to the one I go on to develop. I needn't insist that such a process be entirely explicit, conscious, or deliberate, but it's hard to see how thought experiments work without some such process.

It's also worth mentioning that it is certainly possible, and almost certainly true, that different people come to learn that knowledge is not justified true belief in different ways; we needn't even involve any thought experiments. One might do the whole thing schematically. Here is one way that might go:

(4) It's possible to have a justified false belief.

(5) It's possible to competently deduce truths from justified falsehoods.

(6) Beliefs competently deduced from justified beliefs are justified.

(7) Beliefs deduced from falsehoods are not knowledge.

So, (3) It's possible to have NKJTB.

Perhaps someone could come to knowledge of (5) in this way. However, as Williamson

(2007, 181-82) points out, the general principles invoked in (6) and (7) are much more difficult to defend than the particular judgment of (2_{cf}). Furthermore, this sort of schematic abstraction away from thought experiments probably does not enjoy suitable generality to have sufficiently broad upshots throughout philosophy; it is not plausible that whenever we attain knowledge via thought experiments, we could just as well have run through a corresponding schematic argument. We do not always, or even very often, know, prior to engagement with a thought experiment, what general principles apply.

So even if the kind of structures that Williamson and I are examining are not essential to knowledge of the Gettier conclusion, the project remains: we wish to articulate a particular sort of way that many people come to learn that knowledge is not justified true belief.

5. Truth in Fiction

Let's return to Williamson's formulation, which I've argued to fail. Williamson gives us the predicate GC(x, *p*): *x stands to p as given in the text of the Gettier story*. The way he treats it, GC(x, *p*) holds any time that x stands to *p* in a way that makes each sentence of the text true, if the text is about x and *p*. But there are more resources available here. I think that what is missing from Williamson's analysis is the recognition of the difference between the literal content of the text, and the content of the story generated by the text. There is more to a story than the literal claims of the sentences used to tell it. In particular, we can invoke the notion of *truth in fiction*.[6] Some, but not all, fictional truths are explicitly stated in the text. It is true in my fiction that Professor McStory was teaching an epistemology course; this is so because my text included the sentence,

[6] The seminal piece on truth in fiction is (Lewis, 1978). See also (Walton, 1990) and (Currie, 1990). For recent developments in truth in fiction, see (Hanley, 2004) and the papers cited therein.

‘Professor McStory was teaching an epistemology course.’ Other fictional truths are not so directly stipulated. It’s true in my fiction that Brian had exactly two eyes, even though I didn’t say so. It’s true in my fiction that Professor McStory was not a wizard, even though it’s consistent with the text I wrote that he was. The challenge of a theory of truth in fiction is to explain all of this. I do not here offer a theory of truth in fiction; I merely invoke the useful concept.

Truth in fiction may appear to be just what we need to rule out the bad Gettier cases. Although it’s consistent with McStory’s text that Joe was in a bad Gettier case, it is *true in the fiction* that Joe was in a good Gettier case. (Compare: although it’s consistent with my text that while relating the story about Joe, McStory grew a third eye, it is true in the fiction that he did not do so.) So we might try using truth in fiction in a reformulation of the Gettier argument.

The ‘first stab’ formulation from §2 was unsound; (2_n) is false, because there are bad stories that fit the Gettier text. But suppose we fix the fictional truths themselves—not merely the text of the fiction. Let GC_{tf} (x, *p*) stand for *x stands to p in the relation in which it is true in the fiction that Joe stands to <I have fifteen minutes>*. Then we have:

(1_{tf}) $\Diamond\exists x\exists p\ GC_{tf}(x, p)$

(2_{tf}) $\Box\forall x\forall p\ [GC_{tf}(x, p) \supset (JTB(x, p)\ \&\ {\sim}K(x, p))]$

Therefore,

(3) $\Diamond\exists x\exists p\ (JTB(x, p)\ \&\ {\sim}K(x, p))$

On this formulation, both premises invoke truth in fiction: it’s possible for a person and a proposition to be in the relation that it’s true in the fiction that Joe is in to his proposition. And necessarily, any such person has NKJTB. This formulation is sound, and we need no counterfactual; (2_{tf}) is a necessary universal, just as we wanted it to be. So is this a satisfying account of the Gettier argument?

Not quite, for at least three reasons. First, on this suggestion, the Gettier intuition stands on the notion of truth in fiction. One might worry, then, that everything would depend on the correct theory of truth in fiction; the Gettier intuition would seem to be *a priori* only if we can access fictional truths, given texts, *a priori*; and this is far from clear. Indeed, on one of David Lewis's (1978) views a truth in fiction claim *is* a counterfactual claim: *p is true in the fiction* is analyzed as, *were the text true, p would be true*. This is (paraphrased) Lewis's *Analysis 1*, the simplest of several theories of truth in fiction he puts forward. On *Analysis 1*, the formulation in terms of (2_{tf}) would then be similar to Williamson's, with respect to the epistemic status of the Gettier intuition.[7] Of course, there are good reasons to think that the Lewis view is not right, and some of them locate the problem just with the too-contingent nature of the counterfactual.

Indeed, one of the best sorts of objection to this view is raised by Lewis himself (1978, 274): 'Suppose I write a story about the dragon Scrulch, a beautiful princess, a bold knight, and what not. It is a perfectly typical instance of its stylized genre, except that I never say that Scrulch breathes fire. Does he nevertheless breathe fire in my story?' It's not true in the actual world that if there were a creature matching Scrulch's textual description, it would breathe fire, but it's true in the fiction that Scrulch does, so the counterfactual account of truth in fiction is wrong. (Because of the possibility that it is a conceptual truth that dragons breathe fire, Lewis goes on to further stipulate that the word 'dragon' isn't used in the text of the story.) Currie (1990, 62-70) gives more arguments

[7] But the views would still not be identical. Consider the modal status of the propositional content of the Gettier intuition. On the Lewis account of truth in fiction, the account above would imply that in order to reach the Gettier conclusion, one relies on knowing (q*): *necessarily, if the fictional truths (for this actual fiction) were true, Joe would have NKJTB*. On Williamson's account, one relies on knowing (2_{cf}): *if some x were to stand to some p as in the Gettier text, then anyone who stood to any proposition in a way matching the text would have NKJTB*. Williamson's (2_{cf}) is false in possible worlds where nearby Gettier worlds include bad cases. But q* may be true in all worlds, as 'actually' rigidifies the embedded counterfactual. A given story has the same fictional truths in every possible world; a given text picks out different fictional truths depending upon the circumstances in which it is told.

against Lewis's account in this vein.

So theories tying truth in fiction too closely to counterfactuals are subject to counterexample. Nevertheless, the correct theory of truth in fiction will very likely share this feature with Lewis's: it will have it that fictional truths cannot be known *a priori*. If we cannot know fictional truths *a priori*, this formulation is not one that obviously meets the relevant criteria.

At an earlier stage of my thought about this material, I defended a view related to the one explored in this section. That account was: (1) It's possible for someone to be in a position like Joe's position in this fiction I'm actually engaging with. (2) Necessarily, anyone in a position like Joe's position in this fiction I'm actually engaging with would have NKJTB. (3) Therefore, it's possible to have NKJTB. The rigidification of the fiction was to allow for us to misinterpret authors and engage with our own private Gettier fictions, which would underwrite our intuitions. On this view, the Gettier conclusion here will be *a priori* if one knows *a priori* that Joe's position in the fiction one's actually engaging with comes to justified belief without knowledge. If—and only if—introspective knowledge is *a priori*, this is plausible. My considered view, below, does not assume that introspective knowledge is *a priori*.

A second problem for the GC_{tf} formulation involves the invocation of *truth in fiction* in the content of the Gettier intuition. Many philosophers of fiction think that our ordinary sentences about fictional characters include an elliptical 'it is true in the fiction that' operator, so it is perhaps not terribly worrying that our Gettier intuitions don't *feel* like judgments about fictions. (See, for instance, Currie, 1991; Walton, 1990) Nevertheless, it seems wrong to suppose that invocation of fiction is an essential part of Gettier reasoning. Non-fictional Gettier cases work just as well as fictional ones to set up the Gettier conclusion. If McStory had presented his story to his class as fact rather than as fiction, they would have gone through the Gettier reasoning in just the same way. So it does not seem as though the concept *truth in fiction* can play a role at this level.

A third challenge: I have presented the argument using the GC_{tf} relation, the *relation in which it is true in the fiction that Joe stands to <I have fifteen minutes>.* Which relation, however, is *the* relation? Joe stands in many relations to the proposition in question. Which of those is the GC_{tf} relation? Specifying one is difficult, but without specifying one, I have not fully offered a formulation.

6. From Truth in Fiction to Entertaining Propositions

There is something attractive about the GC_{tf} formulation given above, but it will not do as it stands. A better application of truth in fiction in a Gettier formulation, I suggest, will make use of the notion less directly. On this view, the fictions that are thought experiments are useful for *picking out* and *thinking about* propositions that are key to the Gettier argument. The person listening to or reading a thought experiment can consider *the set of propositions that are true in the fiction*, and in particular, the proposition that every member of that set is true, and then subsequently reason with that proposition to the conclusion of the thought experiment.

I have been assuming that it makes sense to speak of the set of fictional truths; this will be so on most, but perhaps not all, theories of truth in fiction. We might worry about Gregory Currie's (1990) account, on which truth in fiction comes in degrees. Perhaps on this account, only an arbitrary distinction could classify all propositions according to a binary 'true in the fiction' or not predicate. For reasons that should emerge below, this possibility does not threaten my view. Fiction is a useful heuristic, not an essential element.

The process would work like this. Begin with the set of fictional truths. Given a fiction, normal speakers recognize that some propositions are true in the fiction, and others are not. So, when Brian hears McStory's story, he may consider the set of propositions that are true in the fiction. He may refer to that set of propositions demonstratively, or by giving it a name—'STORY', say. Now, Brian may go on to entertain *g*, the proposition that *every element of that set is true,* or *every element of* STORY *is true*. This proposition will play the crucial role in the Gettier reasoning. Now, Brian may reason thus:

(1_g) $\Diamond g$

(2_g) $\Box (g \supset \exists x$ (x has NKJTB))

Therefore,

(3) $\Diamond \exists x$ (x has NKJTB).

Here, the invocation of fictional truth explains how we come to entertain the proposition *p*, but the concept FICTION does not enter into the Gettier reasoning itself. Competence with truth in fiction is an important step in engaging with the thought experiment, but its role is exhausted before the actual invocation in reasoning of (2_g), the Gettier intuition. Put another way, one's ability to grasp a story told through a text serves only as a *means to grasp* a certain proposition, which will figure into the intuition; the content of the intuition itself makes no use of the notion of truth in fiction. Of course, the ability to grasp stories through texts involves (perhaps tacit) *a posteriori* knowledge, albeit knowledge that even (normal) small children possess to some degree.[8] However, this *a posteriori* knowledge, tacit or not, does not prevent the thought-experiment reasoning from being *a priori*. Consider an analogy. Suppose that Professor McStory utters the

[8] For some psychological studies on the way humans think about fictional worlds, see (Skolnick & Bloom, 2006a; 2006b). Weisberg (*née* Skolnick) and Bloom are currently working on a more directly-related question: how children and adults come to recognize fictional truths that are not explicitly stated. Results are not yet published.

following sequence of sentences:

> (i) If Julius Caesar's favorite number is the successor of the number one, then Julius Caesar's favorite number is identical to the number two.
>
> (ii) Two is a prime number.
>
> Hence,
>
> (iii) If Julius Caesar's favorite number is the successor of the number one, then Julius Caesar's favorite number is a prime number.

For his students to understand Professor McStory's verbal reasoning, they need *a posteriori* knowledge: they need whatever *a posteriori* knowledge is required for interpretation of these sentences. The students' *a posteriori* knowledge of English allows them to come to grasp the propositions *if Caesar's favorite number is the successor of the number one, then Caesar's favorite number is identical to the number two*, *two is a prime number*, and *if Caesar's favorite number is the successor of the number one, then Caesar's favorite number is a prime number* from his English utterances, which, in turn, leads them to run through that bit of reasoning. The Swahili exchange student in McStory's class may not have sufficient *a posteriori* knowledge of English; as a result he may fail to run through the reasoning. The reasoning, however, is not *a posteriori*, because the *a posteriori* knowledge the students deploy is not deployed in the content of their reasoning, but rather as a means to get to that content. The reasoning is *a priori* because of the *a priori* status of the premises and the *a priori* entitlement to move from the premises to the conclusion.

Likewise, even though McStory's students need competence with fictions in order to grasp the relevant propositions for the Gettier reasoning, this does not imply that the knowledge they come to receive is *a posteriori*. Only if the *a posteriori* knowledge were deployed as *warranting* could we so conclude. But it is not so deployed. It is merely deployed *as a means to come to grasp the propositions involved in reasoning involving* *the thought-experiment intuition*. The *a posteriori* knowledge serves as a sufficient (not a necessary) *causal enabler* for the reasoning process; it does not play a warranting role within the reasoning itself.

Here is an example. Suppose that one of Professor McStory's students, David, is linguistically incompetent with respect to truth in fiction. So he lacks the *a posteriori* knowledge we typically rely on to engage with fictions. In many cases, the stories he generates from texts differ radically from those that most people generate from the same texts. If we suppose, however, that on the occasion that Professor McStory relates the Gettier text, David generates, merely by luck, the Gettier story most people generate, David may come to the Gettier conclusion via the reasoning given above—this, in spite of the fact that David does not have even tacit knowledge of the principles of generation for truth in fiction. It makes no difference whether David really is in touch with truth in fiction, or even whether he considers the text to be a presentation of a work of fiction at all; if he treated the text as assertive testimony, he could still reason in just the way we describe. That people regularly come to have a (tacit or explicit) representation of the same story explains how it is that we can so easily *share* thought-experiment intuitions; but it is inessential in explaining how some *particular* person comes to a conclusion on the basis of a thought-experiment intuition on a particular occasion.

7. The Psychology of Thought-Experiment Intuitions

One might object to this formulation of the Gettier reasoning on phenomenological grounds. Does it really *seem* that people go about naming the collections of propositions constituting a story as a precursor to engaging in the Gettier reasoning as we have suggested? Probably not. Nonetheless, it does seem that in many cases the Gettier intuition may come to many people with a demonstrative in the content; the Gettier intuition might come to many as:

(2_{that}) Necessarily, if things are like *that*, then someone has NKJTB.

where 'that' refers to how things are according to Professor McStory's story. As demonstratives in a context and proper names share many of the same characteristics, it's not too far of a stretch to think of demonstratives as functioning in a way relevantly (for my purposes) like proper names. If the Gettier intuition comes as something like (2_{that}), then, as with (2_g), the concept of truth in fiction does not enter into the content of the Gettier intuition. Rather, one's competence with truth in fiction goes towards fixing and apprehending the reference of the demonstrative. Thinking of the demonstrative as a proper name, it's not too hard to see that in many cases something very much like the account given in the last section may actually be the reasoning process people go through when they run into Gettier cases. In these situations, people aren't explicitly baptizing with a proper name, but the tacit reference-fixing of a demonstrative amounts to something importantly—and, for my purposes, sufficiently—similar.

We might also wonder what relation it is that we hold to the set STORY when considering the proposition *g* for the purpose of Gettier reasoning. That we can refer to the set allows us to entertain the thought *g*, but it is obvious that we must be better-acquainted with the specifics of the set than mere reference guarantees, in order for us plausibly to know either premise of my formalization of the Gettier argument. I do mean for the person going through the Gettier reasoning to have the Gettier scenario—the set of fictional truths, which goes beyond the literal claims of the Gettier story—*in mind*. (I am understanding the scenario and the set of propositions to be identical.) This needn't require that the subject examine each proposition of the (perhaps infinite) set individually. What is it to have a scenario in mind? It is simply to represent the world as being in a certain way. Someone may have in mind that she receives a major promotion.

Perhaps she is imagining or desiring that scenario to obtain. She could never hope to state a long conjunction of features of this scenario that would entail it; nevertheless, she has this scenario in mind. Indeed, there could be features of her scenario to which she has never explicitly attended. Perhaps she describes the scenario to a friend, and the next day she receives a promotion; everything she said to her friend about her scenario has come true. Nevertheless, it might not be that the scenario she had in mind has been realized; perhaps some feature of it, which she'd never consciously observed, has failed to be present in the real-world version. ('In the scenario I had in mind, I was promoted into a new position; but the way it really happened was that I was promoted because my good friend was fired.') This is all to motivate the idea that we can have scenarios in mind that are constituted by sets of more propositions than we consciously entertain. (A more plausible restriction would be, for any proposition, if it is a part of the scenario, the person with the scenario in mind who sufficiently understands the proposition could, given sufficient introspective skill, adjudicate whether that proposition is a part of the scenario.)

Note that having a scenario in mind needn't be having a possible world in mind; that there are some propositions that are true in the scenario without ever being noticed does not imply that every proposition or its negation is part of the scenario. In the case above, it was part of the scenario that the subject's friend didn't get fired; that there is a French embassy in Thailand is plausibly neither entailed nor excluded by the scenario.

So I don't think there is reason to worry here; among the mental representations frequent among humans are representations of fairly complex scenarios, with elements that needn't be explicitly considered. This is the kind of representation I have in mind when I say that the subject considers the scenario where the fictional truths obtain, then the proposition *g*, that that scenario obtains.

Of course, the question of how well my account—or any other, for that matter—matches what actually goes on in people's heads ultimately must be answered through empirical investigation. While it may be a virtue of my account over one like Williamson's that it better matches the way people reason with thought experiments, I don't want to rest too much on that point. I'm interested primarily in establishing the plausibility of this claim: people could have *a priori* justification for believing the conclusions traditional philosophers often take thought experiments to substantiate, by going through a process similar to the one that many people actually go through. (I have not yet argued that we could have *a priori* justification on the formulation given above; this is coming in the following sections.) Whether it is the case that people are actually *a priori* justified in believing such conclusions I leave, to some extent, as an open question whose answer could only be determined by a more thorough examination of how people reason, and how similar that kind of reasoning is to the ideal practices that would yield *a priori* justification.

Let us now turn to the epistemic status of the premises of my version of the Gettier argument, beginning with premise (1_g).

8. Apriority and Possibility

The first premise in my formalization of the Gettier argument is:

(1_g) $\Diamond g$

First, a minor modification; this first premise isn't *quite* right as it stands. The proposition *g* includes facts about Joe. Joe is a fictional character, and as Kripke says, he may be essentially so. (Although it is possible for there to be some person that matches the description of Sherlock Holmes, we might doubt that any such possible individual would *be* Sherlock Holmes.) If this is right, then it is not possible for all of the members of

STORY to be true, for it is not possible for any person to be Joe. The solution to this worry is technical and peripheral to the larger project; I therefore gave the simpler, more intuitive version above, and will return to it after flagging the issue here. Although it is not strictly possible for fictions involving characters picked out by proper names (who are not inhabitants of the actual world) to be true, there are available modified versions of the fictions which are possible. Professor McStory's story began with the metaphysically impossible, 'Joe had left his watch at home.' A modified, metaphysically possible version of the story would insert, at the start of the story, something like the computer scientist's variable declaration: existential introductions to all the elements in the story that would have had proper names, along with instructions as to how to refer to them while engaging with the story. So we'd have something like: 'There was some guy, whom we will refer to as 'CHARACTER1'. CHARACTER1's name was 'Joe'. Joe had left his watch at home….' If every object with a proper name is introduced thus, then the story is metaphysically possible.

On the more technical account, then, we replace the story with a new story whose truths describe possible worlds that are qualitative duplicates to those described in the old story, taking care to introduce each character (or other proper-named object) in the way described. Then we can reason as described above, using the set of truths in the new fiction, STORY′, and the proposition *g*′, that all the members of that set are true. So understood, the first premise is true. But how do we know it?

In order for the Gettier argument to confer knowledge in the Gettier conclusion, we must be confident that we can know (1_g). And in order for this knowledge to be *a priori*, as I'm attempting to establish, we must be able to know this premise *a priori*. But there are worries about *a priori* knowledge of possibility; Kripke (1980) famously showed that some propositions, such as *Hesperus is closer than Phosphorus*, are *a posteriori* impossible. The modal facts in the neighborhood require empirical investigation to establish. Since there are *a posteriori* necessities, we cannot know modal truths without empirical investigation, lest their negations be *a posteriori* necessities. So goes a common worry. See, for instance, (Yablo, 1993) for an articulation.

I do not think that these skeptical worries pose any great threat to my (1_g). To be sure, Kripkean considerations do demonstrate that sometimes empirical knowledge rules out the possibility of propositions we thought *a priori* to be possible, as in the case of *Hesperus is closer than Phosphorus*. But broad skepticism about possibility does not seem to be in order. There is a useful notion of *conceptual* possibility that can play a role here. In this section, I'll sketch out an account of modal epistemology that I find plausible; chapter four is a more thorough presentation and defense of this view.

It is conceivable, or imaginable, that the members of STORY are all true. Imaginability is often thought to be a guide to possibility; but philosophers defending a modal epistemology along these lines need some way to account for the Kripke-style worries. (See, for instance, Bealer, 2002; Chalmers, 2002; Yablo, 1993) I agree with the critics who argue that Kripkean considerations support skepticism about the idea that this kind of conceivability is generally a good guide to metaphysical possibility. Nevertheless, there does seem to be a useful notion of *conceptual* possibility to which this conceivability is an excellent guide. Conceptual possibility is closely tied to what one can rationally and coherently conceive. If a proposition is a conceptual possibility, then an ideal rational agent can coherently conceive of it as true, and hence have no reason to conclude *a priori* that it is false. Understood as a claim of conceptual possibility, then, premise (1_g) should not be controversial. While there may be some reason to be skeptical about conceptual possibility—as we are not ideal rational agents, and we know that we are sometimes susceptible to paradoxes—its close ties to our own cognitive abilities, which approximate ideal rationality at least some of the time, make radical skepticism with regards to it implausible. That we are fallible detectors of incoherence does not mean that we cannot know sets of propositions to be coherent.

(Some philosophers object to the term ‘conceptual possibility’ on the grounds that propositions like *Hesperus is not Phosphorus* ought not to be judged possible *in any sense*. (See Bealer, 2002; Jackson, 1998) As far as I can see, this is a mere terminological disagreement; those hating the idea of something weaker than metaphysical possibility traveling under the name ‘conceptual possibility’ are invited to substitute their own preferred term.)

But perhaps, given the infinite set that (1_g) is about, our skeptic will not be reassured. After all, one of the occasions in which we generally fall short of ideal rationality is when large amounts of computation are required. Yet, surely large amounts of computation are required to check that all of the propositions true in the Gettier story are coherently co-maintained and consequently conceptually co-possible. As such checking would seem to be prerequisite to knowing that premise (1_g), understood as a claim about conceptual possibility, is true. So it would seem that our rational failings prevent us from knowing as much.

So, how can we limited agents know that the members of the infinite set of propositions true in the Gettier story could be coherently maintained by an ideal rational agent? This worry is answered by reference to a feature of the way that we fix fictional truths. Fictional truths are, at least typically, generated so as to maintain collective coherency. The correct principles of generation for truth in fiction—whatever they are—will have this feature: they will not take us from coherent sets of interpreted sentences to incoherent sets of fictional truths. Fictional truths are what an ideal rational agent with our background knowledge and other cognitive abilities would judge them to be given the interpreted Gettier text. Given the coherency constraint on generating fictional truths, there's no reason to be especially skeptical as to whether the Gettier story is conceptually possible. One's grasp of the proposition *g* comes with an understanding that the propositions it entails are collectively coherent.

This should settle the matter if (1_g) is read as a claim about conceptual possibility. Of course, if this is so, then the conclusion of the Gettier argument, if it is to be valid, can also be at best a claim of conceptual possibility—that it is conceptually possible for there to be NKJTB. Whether this is satisfactory will depend on broader commitments about philosophical methodology. Philosophers like Alvin Goldman who believe the targets of inquiries such as this one to be psychological concepts (2007) will rest satisfied at this point. But many philosophers will insist on a stronger conclusion—they want to establish that NKJTB is genuinely possible—*metaphysically* possible. This is the project I claimed for myself in the introduction. We insist on a metaphysically possible world in which there is NKJTB. To establish the Gettier reasoning, then, I need *g* to be metaphysically possible. Thus, one might object that although we can know *a priori* that *g* is *conceptually* possible, it could only be known *a posteriori* that *g* is *metaphysically* possible. After all, we know from thought experiments from Kripke and Putnam that sometimes, coherently conceivable scenarios turn out to be metaphysically impossible—something that requires empirical investigation to establish.

This worry demands a more thorough examination of modal epistemology than I will pursue in this chapter. However, the strategy for a modal epistemology that I will pursue in chapter four should make it clear that we can establish the Gettier scenario to be metaphysically possible, and not merely conceptually possible, and indeed that we may do so from the armchair. So even philosophers like myself who are interested in establishing the metaphysical possibility of NKJTB should not have reason to object to the epistemic status of premise (1_g). We may still hope, in Kornblith's (2003) terminology, to 'investigate knowledge itself'—even from the armchair.

9. Conceptual Role, Meaning, and Justification

What about my gloss on the Gettier intuition?

(2_g) $\Box (g \supset \exists x$ (x has NKJTB))

How plausible is it that we know (2_g), and that we know it *a priori*? The main argument will be this: naturalistic skepticism about (2_g) is unwarranted; that which explains our capacity for everyday knowledge (whatever that is) should also be able to explain knowledge of thought-experiment intuitions. Consequently, that there is nothing mysterious about the invocation of thought-experiment intuitions. Since there must be an explanation of everyday knowledge, there is an explanation of our knowledge of thought-experiment intuitions. The argument begins with two basic truisms of metasemantics.

The first truism is this: meaningful terms have proper uses. There can be no meaningful term that doesn't have some way that it is used. We may call the proper use of a concept its 'conceptual role'. But this is not to commit to a robust conceptual role semantics—we might disagree about how tight a connection there is between meaning and conceptual role—but we should all agree to the modest Wittgensteinian point that all meaningful terms have proper and improper uses.

The second truism is this: If we have knowledge with a concept as a constituent, we must have some ability to apply that concept with knowledge, at least in some favorable circumstances. We may illustrate the need for this truism via an amusing brief fiction, adapted from one given by Brian Weatherson, (2004, 4), to make an unrelated point:

Cats and Dogs

> Rhodisland is much like a part of the actual world, but with a surprising difference. Denizens of Rhodisland employ a particular concept, CAT*, in all and only the circumstances in which we employ the concept CAT (i.e. when they want to think about cats), and another concept, DOG*, in all the circumstances we employ DOG. But their concepts do not have the same referents ours do; in fact, CAT* refers not to cats but to dogs; and DOG* refers to cats. None of the Rhodislanders are aware of this, so they frequently think and say false things when asked about cats and dogs. Indeed, no one has ever known that their concepts had these referents, and they would probably investigate just how this came to be in some detail, if they knew it were true.

This story is impossible, and this second truism explains why. For any meaningful concept which is a constituent of any piece of my knowledge, I must be able reliably and accurately to apply that concept, at least in favorable circumstances. There must be some cognitive system that, given sufficient relevant input, can reliably and justifiably return whether or not I am confronted with an instance of that concept.

Note that this is not to commit to the implausible claim that we cannot apply concepts in either incorrect or unjustified ways. For one thing, it is consistent with both truisms that we sometimes carelessly ascribe concepts—the second truism demands that we be *able* to apply them accurately, not that we do so in every occasion. We may even be subject to systematic confusions. Another way we can end up with false or unjustified beliefs is to have false or unjustified beliefs about what the situation is like, and then apply the concept in a way that would be correct if the situation were the way we thought it were. That is to say, it I might falsely or unjustifiably believe someone to be an unmarried man, and ascribe to him the concept BACHELOR, properly executing the conceptual role, but resulting in the false or unjustified belief that he is a bachelor.

The correct metasemantic theory of meanings and concepts must explain both truisms. Any non-skeptic about ordinary knowledge is committed to this claim. Without an explanation of this sort, there will be no explanation of how we have any knowledge at all. However—and this is the key point—any explanation for our ability to track the truth of propositions deploying some concept in our day-to-day life will generalize to an explanation for our ability to do the same with respect to thought experiments—that is, to deliver reliable thought-experiment intuitions. We have a conceptual role associated with the concept KNOWLEDGE, and this cognitive faculty tracks whether the things we encounter are instances of knowledge. Once we explain this conceptual role, there is no mystery left about thought experiments. We must all believe in a cognitive mechanism that takes perceptual and belief inputs and reliably outputs ascriptions (in the form of beliefs) of knowledge and non-knowledge; just this mechanism can and does also take imagination inputs and reliably output ascriptions (in the form of imaginings) of knowledge and non-knowledge. This is just an instance of the familiar point that cognitive inference mechanisms treat isomorphic beliefs and imaginings in parallel ways. (See Nichols, 2004.)

So: how does one track whether Joe has knowledge in the McStory story? One uses the same conceptual role associated with *know* that one can use to track whether people actually know or not. If one actually encountered a case like Joe's, the conceptual role one associates with 'know' would give one the resources to know that it wasn't an instance of knowledge. One uses those same resources when one encounters the case through fiction. (And likewise for the conceptual roles associated with 'justified', 'true', etc.)

Consequently, there's nothing mysterious about thought-experiment intuitions. All concepts have associated conceptual roles. These conceptual roles deliver warranted beliefs about certain cases under favorable conditions. Sometimes we believe these cases to obtain, having encountered them through perceptual faculties. Sometimes we imagine them to obtain, having encountered through thought-experiment fictions. Presumably there's nothing suspiciously mysterious about how conceptual roles deliver knowledge about actual cases; whatever explanation we can give there should work just as well for the non-actual cases presented to us via thought-experiments.

10. Tracking Truth in Near and Distant Worlds

These results shouldn't be surprising, as they are harmonious with much that we should expect for other reasons. It's hard to see how one could recognize correct or incorrect deployment of a concept one possesses in actual cases without being able to recognize its correct or incorrect deployment in nearby non-actual cases. So any explanation that explains our ability to apply the concept correctly in actual cases will also explain our ability to apply the concept in many counterfactual cases. To take advantage of our ability to recognize correct deployment of a concept in non-actual cases, one merely needs a medium through which non-actual cases might be presented. Thought experiments give us such a medium in a convenient way. (There are less convenient alternatives that might work just as well. Instead of telling short fictions, we might lie to subjects, attempting to induce beliefs in these non-actual stories, and inviting their judgments. We prefer thought experiments for obvious reasons.)

The foregoing considerations suggest that *any* thought-experiment intuitions attributing concepts (whether 'conceptual analysis' intuitions or moral intuitions about trolley cases, or even the scientific intuitions of Galileo and Schrödinger) should be accurate so long as, first, our actual faculties for recognizing correct or incorrect deployment of the relevant concepts are genuinely good, and second, the counterfactual cases under consideration are relevantly similar to what we might encounter in the actual world. Hence, we have some reason to think that any thought-experiment intuition should allow for knowledge so long as we are sufficiently good at deploying the concepts constituting the propositional content of the intuition in actual cases and so long as the thought experiment isn't so very far-fetched. Gaining knowledge through thought experiments isn't mysterious; it's done continuously with the way that we gain knowledge everyday.

Perhaps the critic's response will be to limit the scope of the skeptical worry. 'Yes,' he might admit, 'thought experiments that are sufficiently mundane and similar to our ordinary experience are acceptable. But this does not vindicate very much of ordinary methodology, which often relies on distant possible worlds, for which our intuitions are ill-suited.' Perhaps my critic will admit that some thought experiments, like McStory's, are acceptable, but insist that far-fetched ones have to go.

In fact, I think there is good reason to suppose us reliable about a considerable number of thought-experiment judgments even in distant worlds. Our concepts are too fine-grained for this to be otherwise. Consider an example:

> Imagine a world that is much like our own, except that instead of oceans filled with water, there are great expanses of orange juice. Someone falls into the orange juice and is submerged; his lungs fill up with orange juice, and he suffocates and dies. Does this person drown?

I take it that it is clear that my unlucky character drowns. We should not be skeptical about our application of the concept DROWNS in this case, even though we are discussing a distant world. If someone responds to this thought experiment by insisting that 'the person does not drown,' this would provide some evidence that he has a different concept at work than we do, albeit a similar one that overlaps in most actual cases. Indeed, distant thought experiments such as this one seem to be just the right way to recognize such subtle divergences in concepts. (I discuss this further in Ch. 5 §4.3, pp. 127-28.)

While it is a matter for debate how much our judgments on these matters have to agree in order for us to possess the same concepts, it is clear that the tolerance of departure in such judgments has a limit before one begins to suspect that disagreement stems from possessing different (often related and nearly actually co-extensive) concepts—this even when judgments concern possibilities far distant from those we actually face. If agreement on such judgments did not have to extend into distant conceptually possible worlds, it's hard to see how our concepts could be so finely individuated as they sometimes are.

The simple point is this: the explanation of how we know conceptual truths is thus continuous with an explanation of how we know anything at all. To explain either, we must explain the connection between possessing a concept and using its associated conceptual role rationally to track whether propositions with it as a constituent are true. Any non-skeptic about ordinary knowledge should expect that there is an explanation for how we do conceptual analysis. Skepticism about intuitive applications of concepts to thought-experiments is unwarranted.

11. Knowledge and Necessity

I haven't quite established (2_g) yet. For (2_g) is a claim of necessity:

(2_g) $\Box$ ($g \supset$ Someone has NKJTB)

The suggestion in the last two sections has been that the competence with concepts needed to explain everyday knowledge is sufficient to explain our warrant for thought-experiment intuitions. So, for example, the faculty underwriting our ability to pick out cases of knowledge in everyday circumstances is the same ability at work in picking out cases of knowledge as presented to us through fictions; an ability to recognize that ($g \supset$ Someone has NKJTB). But one might think that there's a special problem with knowing that *necessarily* the proposition is true. What explains our knowledge of necessity?

The thorough answer to this question would be a complete modal epistemology—again, see chapter 4. The shorter answer is simply to point out that there's no special problem here. The concept *necessarily* has associated with it a particular conceptual role, too. Indeed, it is not too difficult to give a rough characterization of that role. If, in imaging an arbitrary possible scenario (i.e. one free of any particular or special features), one concludes that *p* is the case, then one infers *necessarily, p*. From *necessarily, p*, one can conclude that in any possible scenario one can imagine, it is the case that *p*. To give an adequate modal epistemology is to show how this conceptual role tracks necessary truths, i.e. it is to demonstrate how the conceptual role produces justified beliefs about what is necessary.

Moreover, presumably there is an adequate modal epistemology, because presumably we have as much everyday knowledge about necessity as we do about knowing, believing, or justification. It's widely known that necessarily, if something a dog, it is an animal. It's also widely known that it's not necessarily the case that it won't rain, even if the weatherman said it won't. (Of course, different notions of necessity are at work here, but we have parallel conceptual roles at work, the difference in conceptual role being the scenarios we're counting as relevantly possible.) These pieces of common modal knowledge provide ample evidence that there is an explanation of how the conceptual role of *necessarily* produces justified true modal beliefs. That we philosophers have not yet come up with a conclusive explanation is an invitation to a further research program, not a reason for broad skepticism.

12. Intuitions and Apriority

I conclude with a few remarks toward the apriority of at least some thought-experiment intuitions. Even if there is an unproblematic explanation for the reliability of our thought-experiment intuitions, why think that thought-experiment intuitions like the Gettier intuition are justified *a priori*? At least, there seems to be a burden of proof argument in favor of apriority. A proposition fails to be *a priori* if some *a posteriori* warrant is necessary for coming justifiably to believe, or to know, it. I've gone to some length to emphasize that although *a posteriori* knowledge has a role to play in the thought-experiment process, this role is not warranting. So the onus is on the person who denies that the Gettier intuition is *a priori* to identify the *a posteriori* knowledge that plays a warranting role for either premise.

This is, of course, only a burden-of-proof move. The way thoroughly to establish the apriority of thought-experiment intuitions would be to defend some theory of apriority—or at least some sufficient condition—and show that the judgments with which I've identified intuitions meet that criterion. Since there is no consensus theory of apriority, and it is beyond my scope to defend one, I limit the discussion here to two attractive kinds of strategies concerning apriority that have been influential.

One way philosophers have tried to carve up the *a priori*/*a posteriori* distinction is via perceptual or sensory experience. (Kitcher, 1980, 2000) Roughly, one has essentially *a posteriori* justification for believing that *p* if one's warrant for believing that *p* rests at least partly on perceptual or sensory experiences. One has some *a priori* justification for believing that *p* if one has justification that is not *a posteriori*. If the distinction is to be made out on these lines, it's hard to see how one's justification regarding thought-experiment intuitions would be in all cases *a posteriori*. Whatever role perceptual and sensory experience plays in deduction, this role is not warranting, at least with respect the apriority of the resulting beliefs.

Another way philosophers have tried to make the *a priori*/*a posteriori* distinction (Bealer, 1999, 2002)—or at least to establish a sufficient condition for apriority (Sosa, 2002)—makes reference to one's ability to entertain a proposition. One has *a priori* justification for believing that *p* if and only if full grasp of the proposition *p* is sufficient for accepting it (or perhaps having an inclination to accept it) or *p* follows logically from other beliefs for which one has *a priori* justification. If the distinction can be made out in this way, it would seem that many thought-experiment intuitions do indeed qualify for *a priori* justification. It does appear, for instance, that merely entertaining the propositional content given in some intuitions is sufficient for having some inclination to believe them. No full possessor of the concept DROWN will fail to be at least inclined to judge that the character in my story drowns; likewise, I think, for KNOWLEDGE and standard Gettier cases.

13. Conclusion

I have offered a sketch of a defense of traditional philosophical methodology and its invocation of thought experiments. The defense relies on the demonstration that ordinary apprehension of truth-in-fiction can provide the means for grasping and imagining thought-experiment intuitions that are true necessities—just as they have been traditionally conceived. Furthermore, there is every reason to think that the cognitive faculties explaining everyday reliability about concept application—something that all non-skeptics are committed to—will apply to the imagined cases generated by thought experiments. So we have every reason to consider such intuitions to be knowledge. Finally, there is no obvious obstacle to knowing such intuitions *a priori* on the account developed.

Although I criticize Williamson with regard to the specifics of his view, the view defended here fits well into his broader project: that of 'de-mystifying' philosophical practices, grounding them in our ordinary knowledge about the actual world. Williamson writes that '[s]o-called intuitions involve the very same cognitive capacities that we use elsewhere.' (2004, 152) I agree. I do not see, however, that any of this offers compelling reason to abandon a traditional understanding of philosophical methodology that embraces both the apriority of thought-experiments intuitions and their form as judgments of necessity.

Chapter 4
Imagination and Possibility

1. Introduction

The Red Sox won the Series, and I know it; I came to know it by seeing it. But they *needn't* have won—they might have lost, and I know this, too. I cannot *see* that the Red Sox might have lost in the same way that I can see that they won. A central question of modal epistemology is: how can we know non-actual propositions to be possible?

I also know facts about impossibility: although it's possible that the Sox lost, I know that it is impossible that they both won and lost. An attractive explanation for this knowledge is that I can perform an analytic *reductio* on the alleged possibility: Suppose they won and they lost. If they won, they did not lose; so, they did not lose—contradiction. But not everything I know to be impossible yields a contradiction by analytic inference. I know that it is impossible for Hesperus to be closer to the Earth than Phosphorus, or for some water sample to contain no hydrogen. Call these latter kinds of propositions—ones that do not analytically yield contradictions—*coherent* propositions. We know some coherent propositions to be impossible; another central question of modal epistemology is to explain this knowledge.

One traditional answer to these questions invokes *imaginability*. What is possible is what you can imagine; what is impossible is what you cannot. This is Naïve Modal Rationalism (NMR):

> NMR: For any proposition *p*, *p* is possible if you can imagine that *p*, and impossible if you cannot.

Naïve Modal Rationalism is false; it faces at least two serious problems. First, it is plausible that individuals vary in their imaginative capacities, but facts about possibility and necessity do not. The man on the street cannot imagine curved spacetime, but the theoretical physicist can; but curved spacetime not both possible and impossible. Second, impossibilities are straightforwardly imaginable in at least some sense: supposition is a form of imagination, and we are perfectly comfortable supposing impossibilities for the purpose of *reductio*. It is plausible, too, that we imagine absurdities when engaging with absurd fictions.[1]

These preliminary worries suggest quick fixes. In response to the former worry, we may invoke *idealized* imaginability; we will speak not of any individuals' imaginative capacities, but of those of an ideal rational agent. Perhaps I cannot imagine curved spacetime, but only because I'm insufficiently open-minded; a smarter version of me could.

To avoid the latter worry, we can restrict the relevant acts of imagination to *coherent* imagination. A subject coherently imagines that *p* when she imagines that *p*, and there is no analytic inference from *p* to a contradiction. Insofar as analytic inference is an *a priori* matter, a subject can know *a priori*, of a given imagining, whether it is coherent. (More on analytic inference in §3 below.)

These two fixes to NMR together yield Strong Modal Rationalism (SMR):

> SMR: For any proposition *p*, *p* is possible if an ideal rational agent could coherently imagine that *p*, and impossible if she could not.

But SMR is also false. There is nothing incoherent about imagining that *Hesperus is closer to the earth than is Phosphorus* (*h*). An ideal agent could coherently imagine that *h*; nevertheless, *h* is impossible, because *Hesperus is Phosphorus* (*i*). Since *i* is knowable only *a posteriori*, we cannot tell, merely by imagining, whether *h* is possible.

Is modal rationalism therefore untenable, or would another conservative revision

[1] (Gendler, 2000) gives a fiction in which five and seven are unequal to twelve; (Priest, 1999) gives one in which an empty box contains a small statue.

yield a plausible option? My project is to suggest such a conservative revision. The result will be a moderate modal rationalism that explains knowledge of possibility and necessity, including the necessary *a posteriori*. It will also yield a framework in which it is plausible that we may have a considerable amount of *a priori* knowledge of metaphysical modality.

In §2, I defend the rejection of SMR against a common response. In §3, I introduce and defend a notion of conceptual possibility and necessity. In §4, I relate conceptual modality to metaphysical modality, and show how the former can be useful in coming to knowledge of the latter. In §5, I argue that in many cases, this method can yield *a priori* knowledge of metaphysical modality.

2. Imaginable Impossibilities

Some philosophers are not convinced by the argument from the necessary *a posteriori* that SMR must be false. They dispute that it is after all possible to imagine propositions like *h*—the best we can do, they may think, is to imagine some other, possible state of affairs, and to *confuse* that for a state in which *h*. In this section, I argue that this misidentification response (MR) is untenable.

Kripke himself was the first to make the MR to alleged counterexamples to SMR:

> …though we can imagine making a table out of another block of wood or even from ice, identical in appearance with this one, and though we could have put it in this very position in the room, it seems to me that this is *not* to imagine *this* table as made of wood or ice, but rather it is to imagine another table, *resembling* this one in all external details, made of another block of wood, or even of ice. (1980, 114)

The MR is now widespread to the point of orthodoxy. Following is a brief sampling:

> In this sort of case, one might misinterpret the imagined situation as a situation in which S; here, the situation is merely one in which one has evidence for S. (Chalmers, 2002, 153)

> To imagine myself truly believing that Hesperus and Phosphorus were distinct, I

> would have to imagine them being distinct; and that I cannot do, no more than I can imagine Venus's being distinct from Venus (Yablo, 1993, 23)

> In response to questions, she replies that she is imagining a world in which there is a colorless, tasteless liquid that comes out of taps and fills lakes but that is not H_2O. Now we have a possible defeater … it is not unreasonable to suppose that she is really just imagining that something superficially resembling water is not H_2O rather than water itself is not H_2O. (Tye, 1995, 186)

The MR's ubiquity notwithstanding, it is deeply flawed. To insist that it is impossible to imagine metaphysical impossibilities is at odds with the best available philosophical and psychological views about imagination. The kind of imagination that is relevant to modal epistemology is propositional imagination—it therefore takes the same kind of object as does belief. A highly plausible thesis connects propositional imagination to belief:

> HPT1: If it is possible for someone to believe that *p*, it is possible for someone to imagine that *p*.

A second highly plausible thesis maintains that it is possible to believe metaphysical impossibilities, such as *water does not contain hydrogen* or *Hesperus is closer than Phosphorus*:

> HPT2: It is possible for someone to believe metaphysical impossibilities like *water does not contain hydrogen* or *Hesperus is closer than Phosphorus*.

Our two highly plausible theses together undercut the misidentification response, for they entail that it is possible to imagine the impossibilities the subjects represent themselves as imagining, and that SMR is therefore false. Both highly plausible theses are true.

Of the two theses, HPT2 is perhaps the more obvious. Suppose someone, at the advent of modern chemistry, performs an experiment that misleadingly indicates that a water sample contains no hydrogen. He sincerely reports: 'I believe that water contains no hydrogen'—he speaks truly in so reporting, and therefore believes the metaphysical impossibility that water contains no hydrogen. It would be absurd to offer an analogue of the misidentification response, thus: 'you *think* you're believing that water contains no hydrogen, but *actually*, you're believing that some non-water but watery *stuff* contains no hydrogen, and confusing that state with a state in which *water* contains no hydrogen.' Our subject represents a paradigmatic case of a false belief about water.

Examples are easily multiplied. Lois believes the necessary falsehood that Superman is stronger than Clark, not the contingent falsehood that some *other* guy who *looks like* Superman is stronger than Clark. Or, for any *a posteriori* false proposition *p*, someone may believe the necessary falsehood that *actually, p*—necessarily false because facts about the actual world do not change when evaluated at other worlds. Or, prior to reading Kripke, many philosophers believed that *Hesperus could possibly have turned out not to be Phosphorus*; since this possibility claim is false, it is necessarily false by S4 and therefore also an instance of HPT2.[2]

There are also compelling theoretical reasons to accept HPT1. The best philosophical and psychological theorizing about the imagination relates imagination closely to belief. On one plausible approach to imagination, for instance, propositional imaginings are *simulations* of beliefs; when one imagines that *p*, one enters into a state that is in some senses similar to the belief that *p*. Alvin Goldman (2006a, 2006b) is a clear proponent of this approach. As I have emphasized throughout this dissertation, it has been widely recognized that imagination plays many, but not all, of the functional roles of belief—when I imagine something sad, for instance, I feel sad, not unlike how I'd feel if I believed the sad content. If imaginings are simulations of beliefs, then, it would be very odd indeed if some beliefs—the ones with metaphysically impossible contents—could

[2] (Stalnaker, 1984) thinks possible worlds are the objects of belief, and that it is therefore impossible to believe the impossible. If it is impossible to believe the impossible on a possible-worlds approach, so much the worse for possible-worlds approaches. (Suppose that Stalnaker is right, contrary to my professed belief: it is impossible to believe impossibilities. Then when I believe that someone believes an impossibility, I believe the impossible.) (See Sorensen (1996).) See (King, 2007) for an argument that Stalnaker can and should admit that we can believe the impossible.

not be simulated. If someone can believe that water contains hydrogen, someone else can simulate that belief, and thereby imagine a metaphysical impossibility.

According to a leading alternative approach, propositional attitudes like beliefs, desires, and imaginings involve having bits of syntax represented in cognitive ‘boxes’. (Nichols & Stich, 2000) To believe that *p* is to have in one’s belief box a token representing that *p*; to imagine it is to have such a sentence in the imagination box. The denial of HPT1 on this model would amount to the claim that the imagination box admits different sentences than does the belief box. But this does not seem to be true; the mechanisms that regulate the contents of our belief boxes seem to be just the same mechanisms that regulate the contents of our imagination boxes. Certain incoherent sentences are automatically removed from both boxes by a particular cognitive mechanism (the ‘UpDater’, in Nichols and Stich’s terminology)—and this mechanism operates without regard to which box houses the relevant sentences. Indeed, the parallels between belief and imagination vis-à-vis patterns of inference prompt Nichols (2004) to suggest that belief and imagination are ‘in the same code’—by which he means that a wide variety of cognitive mechanisms process beliefs and imaginings in the same ways. A belief is treated in a very similar way to an imagining with the same content. It is implausible, then, that a metaphysically impossible proposition could be represented in the belief box, but not in the imagination box; no appropriate mechanism is sensitive to the difference.

There are compelling reasons to accept both HPT1 and HPT2. So the Misidentification Response is misplaced and SMR is false. It is possible coherently to imagine metaphysical impossibilities.

3. Conceptual Possibility and Necessity

We must admit, then, that an act of imagining does not entail that its object is metaphysically possible. But it does seem as though there is some interesting status in the neighborhood of possibility being picked out by coherent imagining. An alternative to SMR involves considering *conceptual* possibility, to be contrasted with metaphysical possibility, as a state relevant for modal epistemology. Conceptual possibility, on this line, could be established by coherent imagination; the counterexamples to SMR, though metaphysically impossible, can be counted as *conceptually* possible.

What is conceptual possibility? A conceptual possibility is a coherent scenario; it is a situation that the constraints of rationality make room for. Metaphorically, it is a point in the conceptual space of an agent. More precisely, a proposition is conceptually possible just in case its holding in some scenario does permit analytic inference to a contradiction. *Some green things have no color* is conceptually impossible, because of the analytic inference from *x is green* to *x has a color*. *Hesperus is closer to the Earth than Phosphorus* is not conceptually impossible, because the inference from *x is Hesperus* to *x is Phosphorus*, though necessarily truth-preserving, is not analytic.

It is controversial that there is a coherent and useful notion of analyticity—this particularly in light of the apparent fact that very few ordinary concepts are definitional. The challenge for my gloss on conceptual possibility is to articulate an adequate characterization of analytic inference—one that does not rely on problematic assumptions about the nature of concepts. Furthermore, I do not mean to associate myself with the bankrupt tradition of analyticity as anything in the unfortunate ballpark of 'truth by convention'. In this section, I will articulate and defend a novel epistemic conception of analyticity. Let us begin by examining an inadequate characterization in terms of weak apriority (WAP):

> WAP: An inference from Φ to Ψ is analytic just in case a subject is blindly entitled to it—that is, absent defeaters, a subject is epistemically entitled to reason according to the inference, even if he is unable to articulate reasons why it should be thought to be truth-preserving.

It is very plausible that some inferences, including many intuitively analytic ones, are weakly *a priori*; unfortunately for the suggested account, among these are some that are poor candidates for analytic inference. It may be, for instance, that absent evidence to the contrary, it is reasonable blindly to infer that one is not in a Cartesian skeptical scenario, or that testimony is a generally reliable source of information. (See Kitcher 2000 and Field 2000.) But it is possible coherently to imagine that one is a brain in a vat, or surrounded by liars.

Perhaps, then, we should consider analytic inference in terms of strong apriority (SAP):

> SAP: An inference from Φ to Ψ is analytic just in case a subject is entitled to reason according to it, regardless of his experiences.

The problem with the SAP condition is that it is too strong; it is not clear that there are any inferences that are strongly *a priori*. It could very well be rational for a subject to doubt even the most transparent analytic inferences if, for instance, a consensus of well-established apparent experts told him that it was invalid, or if he acquired evidence that he'd ingested a drug that severely impairs rationality.

So we should admit that there are possible cases in which analytic inferences would fail to be knowledge-transferring. We needn't give up, however, on a principled epistemic notion of analytic inference. Analytic inferences are distinct from nonanalytic ones in that, although they sometimes fail to preserve knowledge, this is so only in particular sorts of circumstances.

Suppose a person is confused—he fails to live up to the standards of ideal rationality. Then he may not recognize the analytic inference from Φ to Ψ, and performance of that inference would not transmit knowledge. He needn't be confused in a pejorative sense (although he may be); some analytic truths are too complicated easily to be recognized. So it is that one may know that x is an elliptical equation, and fail to know that x can be correlated with a modular form, even if he performs that (presumably analytic) inference, if he does so irresponsibly, failing to recognize that the inference is analytic. I borrow the example from (Boghossian, 2003), who uses it in a different way. (The Taniyama-Shimura Conjecture, a lemma in the proof of Fermat's last theorem, is that all elliptical equations can be correlated with modular forms.) This is one kind of way in which an analytic inference can fail to transmit knowledge. The counterexamples to SAP provided another: one may have evidence that suggests that one is relevantly failing the standards of rationality. In these cases, one proceeds unreasonably if one performs the inference, and so it is not knowledge-transferring. These two kinds of failures can occur in both analytic and nonanalytic inferences. What is distinctive about analytic inference, I claim, is that these failures exhaust the possible ways for analytic inferences to fail to transfer knowledge (FTK):

> FTK: An inference from Φ to Ψ is analytic just in case the only possible cases in which the inference fails to transfer knowledge are those in which (a) the subject is failing to live up to the standards of rationality, or (b) the subject has evidence suggesting that (a) is the case.

The FTK test is best understood through its applications to examples.

> If one knows that *Stephen knows that p* and infers on this basis that *p*, under what circumstances can he fail to know that *p*? Only if he is failing with regard to rationality, or has evidence to that effect. So this inference is analytic.
>
> If one infers on the basis of no previous premise that *X is self-identical*, under what circumstances can he fail to know it? Again, only if he is insufficiently rational, or has evidence that he is. So this inference is analytic.
>
> If one infers on the basis of no previous premise that *I am not a brain in a vat*, under what circumstances can he fail to know that he is not a brain in a vat? As in previous cases, it fails to constitute knowledge if he is insufficiently rational, or he has evidence that he is. But, even if this proposition is weakly *a priori*, there is

> another way it may in which the inference may fail to deliver knowledge—if the subject is in fact a brain in a vat. So this inference is not analytic. (Moral: all analytic inferences are necessarily truth-preserving.)
>
> If one knows that *Hesperus is a planet* and infers on this basis that *Phosphorus is a planet*, under what circumstances can he fail to know that Phosphorus is a planet? As before, if he is insufficiently rational or has evidence to that effect, but also if he doesn't know that Hesperus is Phosphorus. So this inference is not analytic. (Moral: not all necessarily truth-preserving inferences are analytic.)

The FTK test distinguishes analytic inferences from *a priori* inferences and from necessarily truth-preserving inferences—and it does so without problematic assumptions about the nature of concepts. It picks out an interesting set of inferences, which in turn picks out the interesting status of conceptual possibility: a proposition is conceptually possible just in case the supposition that it holds in some scenario does not permit analytic inference to a contradiction. (Why the requirement that it holds in *some scenario*, instead of its simply being true? For the canonical contingent *a priori*, like *I am here now*. This is not conceptually necessary; there is no incoherence in my supposing that I am somewhere else.)

4. Conceptual Modality and Metaphysical Modality

Conceptual possibility, unlike metaphysical possibility, is closely tied to ideal imaginability. But why should that be of any interest to modal epistemology? Modal epistemology is about objective modal facts, not about concepts.

Broadly speaking, there are two general strategies one might employ to answer this worry. One might maintain that, once the distinction between conceptual possibility and metaphysical possibility is made clear, it is the former that modal epistemology ought, after all, to be concerned with. This suggestion would be supplemented with either the claim that modal epistemology has actually been about conceptual possibility all along, or an argument that we ought to change the subject in this way. I will not take this strategy; on my proposal, the relevant questions in modal epistemology always have been, and should continue to be, questions about metaphysical possibility and necessity.

On this approach, then, conceptual modality is useful as an intermediary step. Knowledge of conceptual possibility and necessity is useful, from the point of view of modal epistemology, because it can help us get to knowledge of metaphysical possibility and necessity.

The challenge, of course, is the necessary *a posteriori*. It is impossible that *Hesperus is closer to the Earth than Phosphorus*—but it is coherently imaginable and hence conceptually possible. It is tempting to conclude that only facts about conceptual modality, and never about metaphysical modality, can be known from the armchair. But this is an overreaction. That some propositions are necessary *a posteriori* entails only that conceptual possibility does not entail metaphysical possibility. But we have more armchair resources than tests of conceptual possibility.

Although some propositions cannot be known *a priori* to be metaphysically possible, others look like plausible candidates:

- *The Red Sox lost the Series.*
- *Sinners are less lucky than non-sinners.*
- *Most tigers are bred in captivity.*
- *Someone has a justified true belief that is not knowledge.*

All of these propositions are conceptually possible; we can know this by coherently imagining them. I mean to defend the intuitive suggestion that we can know them to be metaphysically possible, too—but how can this be so, given the observation that conceptual possibility doesn't entail metaphysical possibility? As in the case above, it is helpful to examine features of the counterexamples to the entailment.

How did we come to realize that water is necessarily H_2O? Kripke and Putnam invited us to imagine that in some far-off world, some non-H_2O substance XYZ has many of the surface properties of water, and asked us to judge whether that faraway substance was water. We replied that it was not. In so doing, we took ourselves to be *committed* to imagining that XYZ was not water. Note that in general, when we question of a scenario whether it is a scenario in which *p*, we have not two but three choices: (1) we answer that it is a scenario in which *p* if we take ourselves to be committed, by imagining that the scenario obtains, to imagining that *p* obtains. (2) we answer that it is not a scenario in which *p* if we take ourselves to be committed, by imagining that the scenario obtains, to imaginings that not-*p* obtains. (3) We reply that it is indeterminate, or we need more information, if we think that imagining that the scenario does not commit us to imagining one way or the other whether or not *p* obtains. (Suppose that the Red Sox had lost the Series. Is this a scenario in which they hit at least one home run?)

Something in the Putnam story, then, *rationally commits* us to imagining that the substance is not water. Plausibly, it is a fact about the actual world: *actually, water is H_2O*. We test the hypothesis by writing this fact out of the scenario under consideration: imagine that in some far-off world, some non-H_2O substance XYZ has many of the same surface properties of water—and imagine also that in the actual world, contrary to actual scientific consensus, water is not H_2O but that same underlying substance XYZ. Under these suppositions, is that faraway substance water? In this case, the clear verdict is 'yes'.

The relevant water-facts are analytically inferable from the H_2O facts (only) in conjunction with the *a posteriori* fact that water is H_2O. This latter fact is a background assumption when engaging with the relevant thought experiment. Far from undercutting the importance of the imagination and conceptual possibility in modal epistemology, close attention to Kripke-Putnam thought experiments actually motivate drawing a close connection; considerations of conceptual necessity play an important role in explaining our judgments about the necessary *a posteriori*.

Generalizing, we can see that the Kripke-Putnam thought experiments effectively point to a connection between conceptual necessity and metaphysical impossibility: if imagining *p* to hold in some scenario commits one to imagining something false about the actual world, then *p* is metaphysically impossible. (FAMI)

FAMI: $\exists(q)[{\sim}\mathrm{A}(q) \;\&\; \Box_c(\mathrm{S}(p) \supset \mathrm{A}(q))] \supset \Box_m({\sim}p)$

Here, '$\Box_c$' means 'it is conceptually necessary that', '$\Box_m$' means 'it is metaphysically necessary that,' '$\mathrm{A}(q)$' means 'the actual world is such that *q*,' and '$\mathrm{S}(p)$' means 'it is true in some scenario that *p*.' The contrapositive of FAMI expresses a necessary condition for metaphysical possibility:

FAMI*: $\Diamond_m(p) \supset {\sim}\exists(q)[{\sim}\mathrm{A}(q) \;\&\; \Box_c(\mathrm{S}(p) \supset \mathrm{A}(q))]$

We now have a necessary condition on metaphysical possibility that is not met by conceptual possibility. This is unsurprising, since we've known all along that conceptual possibility is insufficient for metaphysical possibility—SMR is false. However, I see no reason to think that the right-hand side of FAMI* isn't, in addition to being a necessary condition for *p*'s metaphysical possibility, also a *sufficient* condition, in conjunction with *p*'s conceptual possibility. Call this claim weak modal rationalism (WMR):

WMR: $\Diamond_m(p) \equiv (\Diamond_c(p) \;\&\; {\sim}\exists(q)[{\sim}\mathrm{A}(q) \;\&\; \Box_c(\mathrm{S}(p) \supset \mathrm{A}(q))])$

To deny WMR right-to-left is to assert that there are some propositions that are conceptually possible, but metaphysically impossible, and for which imagining them true does not necessitate imagining anything false about the actual world. This is the claim that metaphysical possibility requires something more than either conceptual possibility or the condition exploited by the Kripke-Putnam thought experiments. What could this mystery ingredient to metaphysical possibility be? The fact that a proposition meets the necessary conditions expressed by the right-hand side of WMR is at least generally thought to settle the question as to whether the proposition is metaphysically possible.

Anyone defending the ‘mystery ingredient’ view must show either that we do make an additional distinction that figures into our conclusions about metaphysical possibility, or defend the view that metaphysical possibility has necessary conditions about which we’re entirely in the dark. Neither alternative looks particularly plausible. I conclude that WMR is true.

5. Modality and Apriority

We have, in WMR, a straightforward methodology for deciding facts of metaphysical possibility: first, check and see that the proposition in question is conceptually possible, then check and see that imagining it doesn’t commit one to imagining anything false to be actual. In what sense does my WMR deserve the name ‘weak modal *rationalism*’?

To be sure, in some cases, facts about metaphysical modality can only be known *a posteriori*. This is a clear upshot of Kripke. However, it is plausible that there are cases of conceptually possible propositions for which the second WMR condition drops out—cases in which imagining that *p* doesn’t commit one to imagining anything at all—and hence nothing false—about the actual world. Imagining that sinners are unlucky seems to be like this. If imagining that sinners are unlucky does not commit me to imagining anything about the actual world, then the second WMR condition—the usually-*a-posteriori* one—is trivially met. Insofar, then, as I can recognize the relevant facts about conceptual possibility *a priori*, I can know *a priori* that it is metaphysically possible that cheeseburgers cause cancer.

To what extent are those facts about coherent imaginability knowable *a priori*? There is less difficulty in supporting the claim that we can know *a priori* that imaginings of certain propositions are incoherent. We can know *a priori* that some proposition is incoherent by finding a *reductio ad absurdum* using analytic inferences that are weakly *a priori*, and hence, that we are entitled to use without any empirical investigation. (At least some deductive/conceptual inferences should qualify as analytic inferences that are also weakly *a priori*.) Whether we are capable of this depends on how smart we are—but we are smart enough sometimes to find them. We are not ideally rational, so we will fall short of finding every available *reductio* of this sort. Nonetheless, while fallible, we are not hopeless.

Indeed, under many circumstances, we relevantly approximate the ideals of rationality. Anyone can see that imaginings of blatant contradictions are incoherent; likewise with imaginings that there are green apples with no color. Other conceptual impossibilities are less transparent, but still recognizable. Some propositions, such as *there is a largest prime number*, or *some set contains all sets that don't contain themselves*, may appear to us mortals at first glance to be perfectly coherent to imagine, but further investigation—the kind that we are capable of, when we put ourselves to it—reveals them not to be.

Prima facie, there is more difficulty in supporting the claim that we can know *a priori* that imaginings of some propositions are coherent. To know as much *a priori* requires knowing *a priori* that there *isn't* a *reductio ad absurdum* for some proposition. Obviously, searching around for one and not finding one may not be the most reliable method of knowing that there isn't a *reductio*; oversight can explain a failed search. But this point is not to be overstated; it is sometimes reasonable to believe that a proposition is coherently imagined when one examines it and finds no contradiction forthcoming, just as it can, in many cases, be reasonable to believe that a house has no people in it when one examines it and finds none.

Fortunately, we have at our disposal an even more effective method of discovering that there is no *reductio ad absurdum* for a particular proposition. One can 'build a model' of a scenario in which the imagined proposition is true using either perceptual simulation or some other more abstract form of representation. The validity of robustly rational inferences guarantees that there is no *reductio* for any imagining we can build a model for. Moreover, building a model seems like how we know that it's coherent to imagine, for instance, that there are blue swan-like creatures. (To do so, we may simulate perceptual representations of blue swan-like creatures; but sensory imagery is inessential here. We could also build a model with invisible swan-like creatures.) Because it's obviously true that there are blue swan-like creatures in the simulated scenario, we conclude that there couldn't be a *reductio* on the imagining that there are blue swan-like creatures. We draw this conclusion *a priori*; no genuine perceptual experience need play a role in warranting the conclusion. Consequently, there's every reason to think that we know *a priori* that the proposition is coherently imagined. The same could be said of many propositions we take to be coherent.

Discovering that imaginings are coherent can't be anymore difficult than discovering that beliefs are coherent, and most of us are pretty confident that most of the beliefs that we have, even if they are false, are not incoherent. Of course, to think that we frequently have *a priori* knowledge of coherence or incoherence is not to think that for every proposition, we know whether it is coherent to imagine it or not, much less that we always know so *a priori*. The status of some propositions is vigorously contested. An ideal rational agent would know whether it is coherent to imagine that some creature has a brain exactly like mine but has no phenomenal experience.

Conclusion

The Kripke-Putnam thought experiments show that possibility can't be so straightforwardly tied to the imagination as traditional strong modal rationalism suggests. Nevertheless, the methodology those thought experiments presuppose accepts that there is still a very tight relationship between possibility and imagination. I have attempted to codify this relationship in WMR; by exploiting weak modal rationalism, we can achieve knowledge of metaphysical possibility, and we can sometimes do so *a priori*.

Chapter 5
Intuitions and Philosophical Methodology

In chapters 3 and 4, I laid out an approach to understanding philosophical invocations of thought experiments. On this approach, when we engage with a thought experiment, we call to mind a fictional scenario, then make a particular judgment of that scenario: that its subject fails to have knowledge, or acts morally wrongly, or is referring to Gödel, etc. A central motivation of the project was to provide a framework in which such intuitions can be thought of without mystery—I was concerned with avoiding the positing of special faculties of philosophical insight, whose purported reliable operation would be difficult to explain.

I have not yet addressed the *experimentalist critique*, which is an extensive recent literature challenging the invocation of intuitions in philosophy in a more direct way. The challenge (one challenge, anyway—the movement is not at all homogenous) comes in the form of a body of empirical work designed to demonstrate that the folk do not always, or often enough, have the intuitions that philosophers say they do. Insofar as traditional philosophy relies on the alleged evidence that the folk have particular intuitions, this traditional methodology is put open to empirical question—and, in some cases, the experimentalist critique presses, it is severely undermined. In this chapter, I explore how this experimentalist critique fares on the conception of thought-experiment methodology I've been defending. If thought experiments work the way I've said they do, what threat, if any, is there in the prospect of survey data that suggest the folk don't have the intuitions we think they do?

Although neither the experimentalist critique nor my response to it centrally involve questions about the imagination, I think that a brief digression into these issues is here justifiable, as it helps situate the view articulated in Part II into the broader philosophical context.

1. Pressing the Experimentalist Critique

In their influential (2001), Jonathan Weinberg, Shaun Nichols, and Stephen Stich articulated the worries that characterized early versions of the experimentalist critique. This paper, which has important roots in Stich (1990), has inspired a broad new movement of experimentalist philosophy. I cannot, in one brief chapter, address the movement comprehensively; instead, I will focus primarily on the worries pressed in this original paper, and highlight a few connections to later work along the way. The goal, as in the previous two chapters, is to vindicate something that is at least in the neighborhood of traditional methodology. I will focus, with Weinberg, Nichols and Stich—hereafter WNS—on questions about the methodology of epistemology; again, much of what I say, and what they say, should generalize.

A central challenge of WNS's worry about projects in the neighborhood of conceptual analysis is that they are ill-equipped to answer normative questions. Traditional epistemology, it is thought, is in the normative business: epistemology concerns what belief-forming faculties we *should* have, or what kinds of beliefs are *good* beliefs, or what kind of cognitive agent is a *rational* one. These are evaluative notions. But, the worry goes, the methodology of traditional epistemology has no resources to provide a satisfactory answer to such normative questions. Consider the thought experiment: we consider some imaginary scenario, then form an intuitive response: *the subject knows*, or *the subject is unjustified in believing*, or *that cognitive state is more epistemically valuable than this one*.

We start getting worried when we realize that intuitions like these are the products of minds that are heavily influenced by idiosyncratic features of our languages and societies.

There are possible individuals and societies who employ different standards of evaluations and have systematically different epistemic intuitions; indeed, tentative evidence is cited that suggests that that there are actual such individuals and societies. Such possible or actual standards are thought to pose a number of challenges to the viability of the traditional epistemic project. In this passage, WNS, citing Stich (1990), lay out the central challenge. 'Intuition Driven Romanticism' is their name for the family of epistemic methodologies that are the target for their critique.

> There might be a group of people who reason and form beliefs in ways that are significantly different from the way we do. Moreover, these people might also have epistemic intuitions that are significantly different from ours. More specifically, they might have epistemic intuitions which, when plugged into your favorite Intuition Driven Romantic black box yield the conclusion that *their* strategies of reasoning and belief formation lead to epistemic states that are rational (or justified, or of the sort that yield genuine knowledge—pick your favorite normative epistemic notion here). If this is right, then it looks like the IDR strategy for answering normative epistemic questions might sanction any of a wide variety of regulative and valuational norms. And that sounds like bad news for an advocate of the IDR strategy, since the strategy doesn't tell us what we really want to know. It doesn't tell us how we should go about the business of forming and revising our beliefs. One might, of course, insist that the normative principles that should be followed are the ones that are generated when we put *our* intuitions into the IDR black box. But it is less than obvious (to put it mildly) how this move could be defended. Why should we privilege our intuitions rather than the intuitions of some other group? (435)

They go on to present evidence suggesting that the hypothetical people they discuss are actual. I will have a bit to say about this data below. First, it is worthwhile to examine an assumption of this experimentalist critique.

<u>2. Intuitions as Evidence?</u>

2.1. The target of the critique is methodology that relies on psychological inputs.

A premise of the experimentalist critique is that traditional epistemology relies centrally on intuitions; that intuitions are, in some sense, the final arbiters is questions about epistemic properties, norms, and values. It is clear that this premise is intended—for without it, the specter of alternate intuitions is not obviously relevant. The proponents of the critique are explicit about this assumption. WNS write:

> The family of strategies that we want to focus on all accord a central role to what we will call *epistemic intuitions*. Thus we will call this family of strategies *Intuition Driven Romanticism* (or IDR). As we use the notion, an epistemic intuition is simply a spontaneous judgment about the epistemic properties of some specific case—a judgment for which the person making the judgment may be able to offer no plausible justification. To count as an Intuition Driven Romantic strategy for discovering or testing epistemic norms, the following three conditions must be satisfied: (i) The strategy must take epistemic intuitions as data or input. (It can also exploit various other sorts of data.)… (432)

It is clear that WNS intend IDR to include a very broad part of traditional epistemology—including, for example, Gettier's famous argument that knowledge is not justified true belief (452).

The assumption that intuitions serve as important evidence in traditional epistemology appears to be an innocuous one; this, presumably, is why it has received so little explicit defense. However, it is worth examining. The most explicit extended defense of the claim that intuitions are used as important evidence in epistemology of which I am aware is chapter 1 of Joel Pust's (2000) book, *Intuitions as Evidence*. Pust argues for the apparently-obvious claim in the obvious way: by pointing to myriad apparent examples of uses of intuitions as evidence in philosophical matters. He cites thought experiments about knowledge, justification, reference, moral rightness, justice, rationality, personal identity, and explanation—in each case, he points to an 'intuitive' judgment that plays a central role in the argument.

2.2. Do philosophers rely on intuitions as evidence?

There is a gap, however, between the observation that many philosophical arguments rely on premises that are 'intuitive' and the WNS assumption that standard philosophical methodology takes facts about intuitions—psychological facts—as evidence. I believe that once this distinction is articulated, we may understand the suggestion that intuitions are important evidence in traditional epistemology, or in philosophy at large, as ambiguous. Although it is plausibly true in one sense, this is not the one that is invoked in the WNS argument. I will state the two senses in question more precisely presently; first, by way of illustration, consider Pust's first example intended to show that intuitions are used as crucial evidence:

> Here is a case (derived from Lehrer …) from that massive literature:
>
> > [I] Nogot's Ford. Suppose your friend Nogot comes over to your house to show you the new Ford automobile he has just purchased. [standard Gettier story omitted] … Do you *know* that a friend of yours owns a Ford?
>
> Most philosophers take the fact that they have the intuition that S does not know that *p* in this case to show that S does not know that *p*. (5)

Pust closes here with a sociological claim: most philosophers take *the fact that they have a particular intuition* to demonstrate a philosophical thesis. He offers no defense of this empirical claim, apparently considering it obvious. It is certainly true that most philosophers take S not to know that *p* in this case; and it is also certainly true that most philosophers have the intuition that S does not know that *p* in this case. But I am inclined to doubt Pust's claim that most philosophers take the fact about their own mental states to *show* that the fact about S is true. Suppose they were asked to defend the judgment about S. The appropriate response would be to cite, for instance, the fact that S's belief that *p* was derived from a falsehood. It is probably true that, upon sufficient questioning, they might exclaim, 'I just have an intuition!' But this is a way of ending the dialectical train of inquiry, not a serious attempt to explicate the evidence.

Very shortly after the passage I quoted above is another bit from Pust that can provide insight into the mainstream diagnosis of intuitions as evidence:

> The analysis of justified belief proceeds in exactly the same fashion. A theory is

> proposed …, and *it is tested by its ability to account for intuitive judgments* regarding the justifiedness or unjustifiedness of particular actual and hypothetical beliefs. That this is so is recognized by many philosophers who reflect on their practice. For example, the epistemologist John Pollock claims that in epistemological analysis:
>
> > [O]ur basic data concerns what inferences we would or would not be permitted to make under various circumstances, real or imaginary. This data concerns individual cases and our task as epistemologists is to construct a general theory that accommodates it.
>
> (5, emphasis in original)

This strikes me as a remarkable passage. Pust cites Pollock as an example of an epistemologist who reflects on his own practice, and recognizes the crucial role that intuitions play in it; yet the quotation Pust selects does not include the word 'intuition' or any of its cognates. Indeed, Pollock is explicit: the basic data is the *acceptability of inferences*. Somehow, from this, Pust takes away the moral that Pollock recognizes that the basic data are intuitive judgments. What has happened?

2.3. 'Intuitions are evidence' is ambiguous.

The answer, I think, is that there is a crucial ambiguity in the suggestion that 'intuitions are evidence', and that Pust has allowed himself to slide between uses. I'll suggest shortly that WNS have done the same.

Take a prototypical case in which intuitions are allegedly important evidence in traditional epistemology, the story that Professor McStory told in my Chapter 3 §1 (p. 46). I consider the thought experiment and end up with the intuition that Joe doesn't know that Mr. Pumbleton thinks he's on time, and go on to conclude that knowledge isn't identical to justified true belief. Pust will say, 'this intuition is a critical piece of evidence for the conclusion of non-identity.' This statement sounds overwhelmingly plausible. However, consider the structure of the Gettier argument, as defended in chapter 3:

(1_g) $\Diamond g$

(2_g) $\Box\,(g \supset \textit{Someone has NKJTB})$

Therefore,

(3) $\Diamond\exists x$ (x has NKJTB).

Recall that *g* is the proposition that the members of a particular set, STORY, whose members are propositions about a guy named Joe, are all true. Neither premise invokes psychological facts about the person running the argument; there is no mention of intuitions at all. I argued in chapter three that the argument is both valid and the argument that is characteristic of coming to know the Gettier conclusion. Is there any space for intuitions to play the role that Pust and WNS assume they do?

Let's back up. What do we mean when we say that 'this intuition is evidence' for a philosophical claim—say, the claim that knowledge isn't identical to justified true belief? We might distinguish between a strong and a weak reading of 'the intuition that *p* is important evidence for *q*.'

Evidence is propositional. According to a strong reading, the intuition that *p* is evidence for *q* just in case the proposition that *I have the intuition that p* is important evidence for *q*. Pust clearly has in mind—at least sometimes—the strong reading. He is sometimes explicit, as when he says, as quoted above, that 'most philosophers take the *fact that they have the intuition that* [*q*] to show that [*q*]'. The strong reading is also explicitly endorsed by WNS, who characterize intuition-driven romanticism as a methodology that takes facts about intuitions as inputs, and generates philosophical theories on the basis of these psychological data.

But there is a weaker reading available for this sentence. On the weaker reading, 'the intuition that *p* is evidence for *q*' means something like the claim that the *intuited proposition*—namely, *p*—is evidence for *q*. It is on the weaker reading that the Pollock quote given by Pust plausibly lends credibility to the claim that Pollock treats intuitions are evidence in his epistemology. *That such-and-such is permissible*—an intuited proposition—is the starting point for Pollock's theorizing.

The weak reading is not an abuse of English. When Holmes investigates a crime scene, there is a perfectly acceptable sense in which he is relying crucially on his beliefs. But this is not to say that facts about his belief states play crucial evidential roles in his theorizing. In typical cases, he reasons with propositions about the crime scene, not his own psychology. He may run an entire investigation with no explicit introspection at all. In a parallel way, we can express a truth by saying that 'all we have to go on is our knowledge'; but 'our knowledge' here refers to something like the set of known propositions—not any collection of our own mental states of knowing, or facts about what knowledge we have. (If, as most epistemologists think, KK fails, then the former set outstrips the latter.) It is on this model that we may understand the weak reading of 'the intuition that *p* is important evidence for *q*'.

Only this weak reading, I think, is plausibly true. There is no obvious place for psychologistic facts to play evidential roles in traditional epistemology. Consider a standard Gettier argument as formalized in chapter 3:

(1) Such-and-such is a possible story.

(2) Necessarily, such-and-such involves a case of justified true belief that is not knowledge.

Therefore,

(3) It is possible for there to be justified true belief without knowledge.

If I came to know (3) via an argument like this one—and like many epistemologists, I did—then, if someone asks me to cite my evidence for (3), then I will cite (1) and (2). I won't mention any facts of the form, *I have the intuition that….* Indeed, it's hard to see how facts of that sort could play an important role in an argument like this one. To insist upon the psychologization of evidence in philosophy is unduly skeptical, just as it is to insist upon the psychologization of evidence in science. In this I fully agree with Williamson (2007, 211).

The methodology I defend as the traditional one, then, is not an instance of intuition-driven romanticism, as defined by WNS.

3. Objections and Replies

3.1. Objection: Premise (2) just is *the Gettier intuition.*

Yes, I'm happy to speak in those terms; that's why I take it to be true, in the weak sense, that an intuition plays an evidential role in this argument. The Gettier intuition is a proposition about Joe, or Jones, or Fords, or fake barns, or whatever; it's not a proposition about my intuitions, or those prevailing in my society. *That I have* the Gettier intuition is irrelevant to the argument. (This is not to deny the obvious truth that, if I didn't have it, I might well fail to be persuaded by the argument.)

3.2. Objection: Intuitions are, in the strong sense, essential evidence for (1) and (2).

This is a substantive claim—and in my view, not a plausible one. But I won't try just now to argue against it. Instead, observe that attempting to saddle these premises with reliance on intuitions is a defensive move; we began with the Gettier argument, which was meant to be a central, prototypical example of the reliance of traditional epistemology on intuition. Upon examining it, we find no invocation of a fact about intuitions as evidence. If intuitions are thought to play an essential role—in the strong sense—in the epistemology of (1) or (2), even if not in what I argued is the paradigmatic Gettier argument, then the onus is on the critic to demonstrate how it is that facts about intuitions play evidential roles.

I doubt this burden can be met. That a story is possible, or involves justified true belief without intuition, are not claims about anyone's intuitions—indeed, further, they seem to be entirely independent of facts about anyone's intuitions. So it's difficult to see why citing someone's intuitions should play any evidential role in establishing either one.

On my own view, we can come to know premise (1) by invoking something in the neighborhood of a conceivability argument, which I articulate in chapter 4. There is no premise here of the form *I am coherently imagining that g*; rather, we know that *g* is coherent simply by competently recognizing it to be coherent (something we do most easily when coherently imagining it)—not by introspecting and discovering that we coherently imagine it, or competently recognize it as coherent. Similarly, premise (2) is known via the application of a quite general competence for concept application, as articulated in my chapter 3. (Similar stories are given in Devitt 2006, Williamson 2003, and Williamson 2007, 188-90.).

3.3. Objection: Traditional epistemologists themselves describe themselves as using intuitions as evidence.

As I've emphasized above, I do think that there is a sense in which it is true that we use intuitions as evidence in cases like the Gettier argument. In much the same way, I think, we describe ourselves as using our knowledge as evidence in general—but the evidence is the propositional content of the knowledge, not the mental state itself. I therefore take myself to be vindicating some philosophical work—including some of my own—that describes itself as centrally involving intuition. If judgments about hypothetical thought experiments are identified with intuitions, then there is an important sense in which intuitions are centrally important to standard methodology. But it is the content of the intuition, not the fact of the intuition itself, that plays a key evidentiary role—this is the entity that is input into the relevant strategy.

I don't mean to deny that some traditional epistemologists have explicitly signed on to the strong interpretation of intuitions as evidence. Nelson Goodman is an oft-cited example—see Stich (1990, 77). Although I'm less than fully convinced that this interpretation of Goodman is correct, I have no desire to press the point. These are subtle matters, and are easily confused. I'm happy to admit that some or even many traditional epistemologists may be on the record as endorsing a conception of traditional analytic epistemology as taking psychological facts about intuitions as evidence. George Bealer (1999, 2002) is certainly an example. I am arguing that such philosophers are incorrect about the nature of their shared project. One does not centrally rely upon intuitions merely by virtue of defending the view that philosophers centrally rely upon intuitions, any more than one fails to know truths about the external world merely by virtue of being an indirect realist. (Let us assume, for the purpose of illustration, that indirect realism implies skepticism.) It is best to look to the methodology itself—not to any particular group of epistemologists' views about the methodology. The experimentalist critique purports to discredit the role that thought experiments play in arguments in traditional epistemology. It is this methodology I mean to be defending; part of my defense involves the observation that at no point do facts about intuitions play a critical role in these arguments.

(Another category of philosophical approaches should be mentioned, if only to be set aside. As I have highlighted in previous chapters, some epistemologists explicitly sign on to a project in which facts about intuitions are important evidence, and do so, I think, without error. I have in mind epistemologists whose project is not the traditional one gestured at above—one about, roughly, the nature and grounds of human knowledge.

Rather, they are concerned with questions about epistemic concepts. Alvin Goldman (2001, 2007) is an exemplar of this approach.[1] The project of articulating the nature and application of epistemic concepts, which has a perfectly obvious and respectable use for psychological facts about intuitions, is not the project that presently concerns me.)

4. Recasting the Experimentalist Critique

If the strong conception of intuitions as evidence were true relied upon by traditional methodology, then the experimentalist critique would appear quite forceful: on a traditional approach, facts about what intuitions you have play a crucial role in determining the correct normative epistemic view. So if other people had importantly different intuitions, then their applications of the standard methodology—even proper applications of it—would lead to divergent views. The challenge of why we should prefer the use of facts about our intuitions to ones about theirs looks serious.

But I've argued that the strong conception is not true; intuitions are only important evidence in the weaker sense articulated above. How does the experimentalist critique fare with only the weak understanding of intuitions as evidence? On the conception of traditional epistemology that I've argued for, facts about intuitions are not typically plugged into WNS's 'IDR black box'. The kinds of propositions that are plugged into the IDR box are propositions like *such-and-such is not a case of knowledge*. Little of traditional epistemology, it seems to me, falls under WNS's heading of intuition driven romanticism.

How much of the experimentalist worry, then, is dispelled, and how much can be recast in terms that do not rely on the premise that traditional epistemology makes critical

[1] WNS (2001) was part of a special issue on the Philosophy of Alvin Goldman; Goldman's (2001) reply to WNS consists largely in the clarification that his work is to be understood as engaging with the mentalistic project discussed here.

use of psychological facts about intuitions as evidence? Certainly, many of the experimentalist critics, when faced with a challenge like the one I've just made, are not much impressed; it is thought that it is not particularly difficult to re-cast their skeptical worries in terms that do not invoke the false assumption. If this is right, then the argument I've attempted to distill from WNS doesn't get to the heart of the experimentalist critique. Three attempts to re-cast the critique without invoking the problematic assumption of traditional reliance of psychological facts as evidence suggest themselves; I discuss them now in turn.

4.1. It is arbitrary and xenophobic to privilege our own concepts and judgments.

The original WNS emphasis was on normativity: how can traditional methodology reliably yield normative conclusions? Perhaps we should understand the experimentalist critique as emphasizing this feature. Sure, we (well-educated, white Westerners who have studied philosophy) value knowledge; but there are *possible* people who value some other, non-knowledge state instead. Here is a challenge: Suppose we successfully articulated what rules we must follow in order for our beliefs to fall under our concept KNOWLEDGE; why should we *care* about following those rules? What value is there in complying with a standard that happens to be reflected in our language and society? It is xenophobic to privilege our own standards merely because they're ours. Stich (1990) writes:

> [U]nless one is inclined toward chauvinism or xenophobia in matters epistemic, it is hard to see why one would much care that a cognitive process one was thinking of invoking (or renouncing) accords with the set of evaluative notions that prevail in the society into which one happened to be born. (94)

And:

> Since our notion of justification is just one member of a large and varied family of concepts of epistemic evaluation, it strikes most people as simply capricious or perverse to have an intrinsic preference for justified beliefs. (95)

The critic pressing this ‘isn’t that nice’ challenge to normative epistemology may well concede that, for example, armchair resources are sufficient for knowledge that the traditional judgment about a Gettier case—that the subject’s state does not fall under the everyday concept KNOWLEDGE—is correct. In so doing, of course, he admits that we know that the Gettier subject does not know. (All parties must agree that S knows that *p* if and only if S’s relation to *p* falls under the concept KNOWLEDGE, and that this fact is easily known by those of us who have the concept.) But the critic isn’t much impressed by this admission. As Stich likes to say: now we know what knowledge is, and *isn’t that nice*? He (1990) writes:

> The analytic epistemologist proposes that our choice between alternative cognitive processes should be guided by the concepts of epistemic evaluation that are “embedded in everyday thought and language.” But this proposal is quite pointless unless we *value* having cognitive states or invoking cognitive processes that accord with these commonsense concepts. And it is my contention that when they view the matter clearly, most people will not find it intrinsically valuable to have cognitive states or to invoke cognitive processes that are sanctioned by the evaluative notions embedded in ordinary language. Nor is there any plausible case to be made in favor of the instrumental value of beliefs or cognitive processes that are justified or rational. (93)

The idea, I take it, is that the interesting questions of epistemology are normative—they’re supposed to help us to know what sorts of beliefs to pursue. Knowing what beliefs fall under our ordinary concept KNOWLEDGE is no help in this normative enterprise unless we have some reason to value having beliefs that fall under that concept; this, Stich says, is implausible.

We must tread carefully here. Here is a fact that I know: the referent of the everyday concept KNOWLEDGE is knowledge; this follows straightforwardly from the fact that I am employing the everyday notion in thinking that thought and writing that sentence. If we keep this in mind, I think it should be clear that the value Stich attributes to the analytic epistemologist—*according with the standards of everyday thought and language*—is optional as the epistemologist's object of value.

Here is a traditional view: knowledge is valuable. The attempt to explain the value of knowledge has occupied considerable attention from epistemologists since Plato. Among the candidate explanations are suggestions like: knowledge is the norm of assertion; knowledge is the norm of action; knowledge helps the subject achieve his interests; knowledge is a more stable kind of true belief; knowledge is part of the Platonic Good; knowledge is a successful achievement of a characteristically human performance.

It is no part of my present project to contribute to this vast literature. The question that concerns me is whether Stich's argument casts doubt on the cogency of the project of treating knowledge as valuable, and seeking the explanation for that value. I agree with Stich that it would be an odd creature indeed who placed great value in the state of *having beliefs that fall under the extension of the everyday concept of knowledge*. Call this state S. Such a valuation is probably not incoherent, but it does appear ill-motivated. It is no great defense of traditional epistemology if it leaves the value of knowledge like that.

Must one value S in order to value knowledge? Plausibly not. Although in fact, in the actual world, all and only people with S have knowledge, S and knowledge are not the same property. Depending on how we individuate concepts, the biconditional that one has knowledge iff one has S may well be only contingently true—there may be worlds where the concept KNOWLEDGE does not refer to knowledge. And there are certainly worlds where the word 'knowledge' does not refer to knowledge.

Etiquette norms (the ones around here) dictate that the fork be set to the left of the plate. Many of us value acting in accordance with those norms. There is, plausibly, at least instrumental value in complying with the rules of etiquette in one's society. The way in which we value setting the fork on the left is contingent on the rule being as it is.

The way we value epistemic norms are different. Knowledge is valuable, regardless of what epistemic ideals happen to be coded into our language. The disanalogy is especially apparent in divergent counterfactuals:

> If our social norm were to put the knife on the right, instead of on the left, there would be no etiquette value to putting it on the left.
>
> If our social norm were to have beliefs that are schmowledge, instead of knowledge, there would be no epistemic value to knowing.

Many of us will accept the first but not the second. That second has some of the feel of:

> If our social norm were to kill all the old people, there would be no moral value to refraining from killing them.

I take it just about everybody who thinks there is actual moral value in refraining from killing old people rejects this one.

Another way to see this point is to observe how far Stich's argument, if sound, would generalize. Take whatever candidate for value that you like—desire-satisfaction, or pleasure, or eudaimonia, or true belief, or whatever you find most plausible. Stich's counterpart can argue:

> You propose that our choice between alternate courses of action should be guided by what falls under our concept PLEASURE. But this proposal is quite useless unless we *value* having states that accord with this commonsense concept. It is my contention that when they view the matter clearly, most people will not find it intrinsically valuable to have states that are sanctioned by the PLEASURE concept that happens to be embedded in ordinary language. Nor is there any plausible case to be made in favor of the instrumental value of states that are pleasurable.

It is a mistake to argue, from the premise that it is implausible to value *matching the ordinary concept of XNESS*, to the conclusion that it is similarly implausible to value xness.

The challenge from arbitrariness and xenophobia takes a similar starting point. Our epistemic evaluations are informed by our epistemic concepts, which are a product of contingent features of our upbringing. WNS write, in Nichols, Stich, & Weinberg (2003), that:

> Without some reason to think that what white, western, high [socioeconomic status] philosophers call ‘knowledge’ is any more valuable, desirable or useful than any of the other commodities that other groups call ‘knowledge’ it is hard to see why we should care if we can’t have it. (245)

Suppose we learned that in some society, the word best translated as ‘knowledge’ carried a different meaning from the ordinary English ‘knowledge’—perhaps their word means justified true belief. (On one interpretation of WNS, their data suggests that certain idiolects of English are like this.) What reason, WNS ask, do we have to prefer our criterion of epistemic evaluation (knowledge) to theirs (JTB)?

Even if it turns out that there are no such societies, a version of this challenge may still be pressed: surely we *could have* used a concept of evaluation like that. So it is in some sense *arbitrary* that we don’t. What reason have we to prefer the notion we happened to end up with to any other? A very plausible response to this challenge seems to me to be the pluralist one suggested by Ernest Sosa (forthcoming): what’s to stop you from valuing JTB? Nothing at all. We value all sorts of things; value whatever you want. This is consistent with continuing to value knowledge. Sosa writes:

> The fact that we value one commodity, called ‘knowledge’ or ‘justification’ among us, is no obstacle to our also valuing a different commodity, valued by some other community under that same label. And it is also compatible with our learning to value that second commodity once we are brought to understand it, even if we previously had no opinion on the matter. (15-16, manuscript)

This response strikes me as entirely correct. But Stich identifies two challenges for this pluralistic line: first, that a satisfactory epistemology should supply norms of permission, not merely norms of valuation—something for which pluralism is less plausible, and second, that even if we limit ourselves to concerns with value, the pluralist has no resources to weigh tradeoffs in value in the inevitable case when one is forced to choose between alternate alleged epistemic goods. On norms of permission, Stich (forthcoming) writes:

> Norms of valuing do play a role in traditional epistemological debates, but they are not the only sorts of norms that epistemologists have considered. As we noted earlier, Goldman insists, quite correctly, that justification rules (or "J-rules") play a central role in both classical and contemporary epistemology, and J-rules specify *norms of permissibility*, not norms of valuing. They "permit or prohibit beliefs, directly or indirectly, as a function of some states, relations, or processes of the cognizer" (Goldman 1986: 60). When we focus on these rules, the sort of pluralism that Sosa suggests is much harder to sustain. If a rule, like the one cited a few paragraphs back, says that *ceteris paribus* we ought to hold a belief if it is an instance of knowledge, and if 'knowledge' is interpreted in different ways by members of different groups, then Sosa's pluralism leads to inconsistency. There will be some beliefs which we ought to believe on one interpretation of 'knowledge' but not on the other. (9, manuscript)

And on trade-offs:

> Moreover, even in the case of norms of valuing Sosa's pluralism can lead to problems. Sosa is surely right to claim that someone who values owning money banks can also value owning river banks. But if there is one of each on offer and the person's resources are limited, she will have to make a choice. Which one does she value more? (9, manuscript)

With regard to norms of permission, it is not at all clear that the fact that some *ceteris paribus* rules will come into conflict with one another is decisive against the relevant sort of pluralism. That they're *ceteris paribus* rules, instead of absolute ones, is just what is required to tolerate this sort of conflict. So if my neighbors use 'knowledge' to pick out JTB, which they value, they and I can all endorse and share the relevant *ceteris paribus* permissibility rules:

> *Ceteris paribus*, believe that *p* if and only if doing so will result in knowledge that *p*.
>
> *Ceteris paribus*, believe that *p* if and only if doing so will result in JTB that *p*.

(Incidentally, it is worth pointing out that it is not clear that there are many actual practical circumstances in which these rules advise an agent in divergent ways; the person who tries to maximize the one will look quite a lot like the person who tries to maximize the other. This suggests that the extent to which genuine conflict among plausible epistemic norms of permissibility may not be as great as Stich assumes. I will not pursue this line of thought further.)

Of course, this defense of pluralism only holds if we treat the norms as *ceteris paribus* rules. Could a version of the critique insist on absolute rules? Not very effectively, for absolute rules of this sort are highly implausible. It is *all things considered* permissible to fail to maximize knowledge in some circumstances, even though doing so will violate some *ceteris paribus* epistemic rules. An all-things-considered knowledge maximization principle would prohibit building houses in favor of counting bricks. As Sosa (forthcoming) points out, this is so even if we limit the relevant domain to the epistemic:

> Silly beliefs about trivial matters can attain the very highest levels of epistemic justification and certain knowledge even if these are not beliefs that one should be bothering with, not even if one's concerns are purely epistemic. (17, manuscript)

So there is no particular difficulty with a pluralist response to the xenophobia challenge that goes along with Stich's preference for *ceteris paribus* norms of permission. And stronger rules are implausible, independent of considerations of alternative norms or values. Stich may well insist: *but which rule should we follow*? This is in effect to assimilate the challenge in terms of rules to the challenge in terms of values formulated above; I will therefore turn to values and consider the challenges together. Stich's second argument is that there is a problem for the pluralist even at the level of value. I say, with Sosa, that there's nothing stopping me from valuing the other societies' epistemic goods in addition to my own, if I can learn to think about them. If some people or societies value true belief, or justified belief, or justified true belief, or belief derived from a generally reliable source, or certainty, but don't even have a word for knowledge, we can all get along just fine, and even learn to value one another's preferred states too.

Stich replies, but *which do you value more*? Since we are finite creatures with finite resources, we must choose among the things we value; the pluralist hasn't told us how to adjudicate between valuable things—and traditional armchair epistemic methodology doesn't obviously have the resources to identify the appropriate criterion. It should be clear that this is exactly analogous to the question just raised about *ceteris paribus* rules.

My answer here is simple: I agree with Stich that traditional armchair epistemology does not obviously provide the resources to adjudicate between different valuable states, or conflicting *ceteris paribus* rules. But I very much doubt it ever pretended to. Suppose we set aside questions about differing epistemic concepts; even if knowledge is the only epistemic game in town, we still have to decide whether to read the encyclopedia or walk the dog. If Stich thinks it is a great scandal that traditional epistemology provides no clear advice on this matter, he has broader expectations for epistemology than I.

I have so far been assuming that the rival epistemic goods, though not identical to our epistemic goods, were not antithetical to them. If we value knowledge, and our neighbors value truth, JTB, or certainty, then there seems to me to be no obstacle to our sharing their values, as explained above. This case to me seems analogous to this one: I like opera. I can get along just fine with Emily who likes Puccini operas, Andrew who likes theatrical performances in general, and even Martin who likes basketball. Not only are we peaceful neighbors, but we can even learn to appreciate one another's particular preferences, and share them to a large extent.

But there could be a person or society with vastly different epistemic values—values that are not only non-identical to ours, but in direct tension with them. Maybe they value false beliefs, or unjustified ones. Or maybe they're Pyrrhonians, who value the complete absence of belief. This is more like the case where I like opera and my neighbor demands total silence—our values just plain conflict. (Of course, there is no evidence on the table that there are people or societies like this.) In this extreme case, I reject the values of my alien neighbors. But this is no xenophobia—I avoid them, not because they are different from me, but because their values are inconsistent with the things I value.

The arbitrariness reformulation of the experimentalist critique, then, need not convince us to abandon traditional methodology.

4.2. Recast the critique by swapping 'judgment' for 'intuition'.

Joshua Alexander and Jonathan Weinberg (2007) provide a response to Timothy Williamson's (2003, 2007) invocation of the view defended in this chapter, that psychological facts about intuitions are not the primary evidence in philosophy. It is worth quoting their response at some length. Alexander and Weinberg write:

> Timothy Williamson has also developed a more radical response to the restrictionist threat: rejecting the picture of philosophical practice as depending on intuitions at all! He argues that our evidence, in considering the cases like those listed in section 1, is not any sort of mental seeming, but the facts in the world. He compares philosophical practice to scientific practice, where we do not take the perceptual seemings of the scientists as our evidence, but the facts about what they observed. Similarly, then, we should construe Gettier's evidence to be not his intellectual seeming that his case is not an instance of knowledge, but rather the modal fact itself that such a case is not an instance of knowledge. We retreat from talk of the world to talk of percepts when we (mistakenly) attempt to accommodate the skeptic; so, too, do we retreat to talk of intuitions only under the pressure of skeptical arguments. And since Williamson is himself antiskeptical, emphasizing the continuity between ordinary modal cognition and philosophical cognition, he concludes we should give up thinking of our philosophical evidence in the thinly psychological terms of intuitions.
>
> But we do not think that Williamson's arguments can provide much solace for traditional analytic philosophers. For the results of experimental philosophers are not themselves framed in terms of intuitions, but in terms of the counterfactual judgments of various subjects under various circumstances. Although the results are often glossed in terms of intuitions to follow standard philosophical usage, inspection of the experimental materials reveals little talk of intuitions and mostly the direct evaluation of claims. The restrictionist challenge does not need to turn on a (potentially mistaken) psychologization of philosophers' evidence; that it does not turn on that skeptical move hopefully helps make clear that it is not itself a skeptical challenge. In terms that Williamson should be happy with, the challenge reveals that at the present time philosophers may just not know what their evidence really is. And the true extent of their evidence is not, we think, something that they will be able to learn from their armchairs. (72)

There is much to appreciate in this thoughtful passage. We must take care, however, not to draw more lessons than are warranted. One might be tempted from this passage to think that the experimentalist critique can be reformulated from a critique of the invocation of 'intuitions' to a critique of the invocation of 'judgments,' and thereby circumvent the response with a simple invocation of the find-and-replace button. But this is in at least some cases (such as the original WNS critique) false. The case I've made, that facts about intuitions are not used as fundamental evidence in philosophy, generalizes against the claim that facts about *judgments* are used as fundamental evidence in philosophy. Williamson's (2003) controversial claim that 'intuition' talk is unhelpful, and that philosophical 'intuitions' are really only particular sorts of judgments, is far from the centrally important observation. The centrally important observation is that philosophical evidence is not primarily psychological at all.

Alexander and Weinberg do have in mind, however, experimentalist arguments beyond the WNS one I've been focusing on. Some of these do point to important reasons for humility in our engagement with sensitive philosophical questions. For example, Swain, Alexander, and Weinberg (2008) have found that when subjects who are presented with thought experiments and asked to judge whether their protagonists have knowledge, their responses seem susceptible to a pernicious order effect: subjects are more likely to attribute knowledge in a tricky case—a fake barn case or a TrueTemp case—if they've recently been asked about an obvious case of non-knowledge than they are if they've just been asked about an obvious case of knowledge. That casual judgment is susceptible to such irrelevant features should not be particularly surprising; there is strong independent reason to believe that humans are susceptible to many such kinds of performance errors (for a nice summary, see Stich (1990), pp. 4-9, citing Nisbett and Borgita (1975), Wason and Johnson-Laird (1970), Tversky and Kahneman (1983), and others). Nevertheless, the point that philosophers must not blindly stick to whatever philosophical intuitions they find themselves attracted to is well-taken. We should be circumspect in our philosophical judgments, especially in situations where we are particularly prone to error. Here, empirical data can surely help us to improve our epistemic positions, by helping us to identify the fallacies and biases to which we are prone. There is important and valuable empirical work to be done in this part of experimental philosophy; in my view, the movement's most impressive achievements occur in this area. For instance, Horowitz (1998) impressively employs Kahneman and Tversky's Prospect Theory, an empirical theory about how humans reason with risks and probability, to discredit certain deontological intuitions in normative ethics. Her project strikes me as entirely compelling. Gendler (2007) offers a catalogue of similar projects, and Gendler (2002) is her own attempt to discredit, on empirical grounds, a particular sort of intuition about personal identity. All of these projects are worthwhile and philosophically relevant; none, I think, undermine the sort of methodology laid out in these chapters on methodology.

Weinberg (2007) expresses pessimism that philosophical judgments are capable of sufficiently careful treatment to figure into a respectable methodology. At present I will say only that this pessimism does not seem to me well-founded; it relies on unfriendly assumptions about the homogeneity of philosophical judgment, and our inability to distinguish shakier intuitions from more solid ones. I see no reason to think that careful philosophers cannot take care to avoid these sorts of errors, the same way that careful perceivers do.

What of my earlier claims, in chapters 3 and 4, that much of philosophical methodology is *a priori*? Does my concession here undermine this claim? Not necessarily. A full defense here would involve a criterion for apriority, which is beyond my present scope, but, as in chapter 3, we may go some way toward establishing the plausibility of the apriority of the methodology by examining proposed criteria for apriority that have been influential. Recall the two alternative approaches to apriority outlined in chapter 3 §12 (pp. 81-82): first, a proposition is known *a priori* just in case, roughly, it is known and one's warrant for it does not derive from sensory experience; and second, a proposition is known *a priori* just in case, roughly, it's known and full understanding of the proposition is sufficient for that knowledge. Both formulations are consistent with claim that *a posteriori* human limitations sometimes lead us to systematic errors, and that we philosophers are well served to attend to such limitations in order to avoid such errors whenever possible. A simple model illustrating this compatibility is one in which the interference of a systematic human error is a defeater for *a priori* knowledge and of full understanding; in the absence of such errors, we are able to achieve *a priori* knowledge. So an empirical understanding of our susceptibility to such errors can help put us in a position to achieve more *a priori* knowledge.

On a similar theme, compare Williamson (2007):

> [P]sychological experiments might in principle reveal levels of human unreliability in proof-checking that would undermine current mathematical practice. To conclude on that basis alone that mathematics should become an experimental discipline would be hopelessly naïve. (7)

I should add that the dialectic is here complicated by the fact that Williamson himself does not consider philosophy to be *a priori*, as he rejects the dichotomy between apriority and aposteriority, primarily on the basis of pessimism about providing an adequate characterization of the alleged distinction (2007, 165-69). So I count Williamson as an ally against Weinberg here only insofar as he agrees with me that much philosophy is legitimately pursued from the armchair. Providing a rigorous characterization of the distinction between apriority and aposteriority is an important project to which I hope to give future attention.

4.3. The experimentalist challenge derives from general challenges about disagreement.

Finally, I am also inclined to be moderately concessive with respect to a third reformulation of the experimentalist critique—one in which the posited divergent intuitions are treated as epistemic challenges to methodology just insofar as our confidence in any particular belief should in general be shaken upon encountering disagreement. So, in particular, discovering people who think that Gettier cases are knowledge undermines traditional philosophical practice because in general, discovering people who believe not-*p* undermines the belief that *p*.

This strikes me as right so far as it goes; however, I doubt it will go very far. A combination of two lines of defense will, I think, serve to vindicate the traditional approach with respect to the clear cases, like the Gettier intuition. First, as Ernest Sosa (2007b) has emphasized, one plausible explanation for apparent divergences in opinion about Gettier cases could easily be that the relevant subjects are not judging with respect to the intended proposition. There are two ways this could be so: first, if they 'filled out' the story of the thought experiment in a nonstandard way (see p. 51-52, about 'bad' Gettier cases, in my chapter 3 §2), and second, if they interpreted key terms, such as 'knowledge', in a nonstandard way; i.e., if they mean something other than (but presumably similar to) knowledge by 'knowledge'. Divergent judgments about the cases in question are exactly what would be predicted by such possibilities. (See p. 78, about drowning in orange juice, in my chapter 3 §10.)

A second mitigating factor against the undermining of apparently-firm philosophical beliefs by disagreement is the possibility that the deviant subjects are ignorant, confused, inattentive, or otherwise cognitively inferior with respect to the judgment that is being asked of them. Of critical importance in how much our beliefs are undermined by disagreement is the authority of our disagreeing interlocutors; if I learn that someone thinks that I am dead, this does not much—if at all—undermine my belief that I am alive. In this matter, I am the relevant expert. (See Elga 2007.) Likewise, the physicist who knows that tables comprise mostly empty space and the astronomer who knows that the universe is billions of years old retain their knowledge, even in the face of the onslaught of folk who disagree with them. Why not so for the epistemologist who knows that Jones doesn't know someone in his office owns a Ford?

There is one obvious difference, of course: the epistemologist's judgment is *a priori*; the physicist's judgment about tables is based on scientific theory and observation, and my judgment about my own life is on the basis of direct experience. People who believe that tables have no empty space inside them, or that I am dead, are lacking important evidence; the folk who think Jones knows are not. So goes one argument.

This disanalogy cannot stand up as presented. As philosophers well know, not all *a priori* investigations are easy; there's no guarantee that just anyone will get these questions right. (Compare the analogous situation when the folk disagree with a mathematician about a mathematical fact.) To recognize Gettier cases as cases of non-knowledge is a cognitive achievement; it is entirely possible that, without philosophical training, some people might fail to achieve it. Happily, there is evidence that philosophical training *does* help—some of WNS's data suggests that people who had studied philosophy were significantly more likely to have the Gettier intuition. (Nichols, Stich & Weinberg 2003, 242.)

It is plausible that a combination of these two responses can defend traditional methodology in these cases to a very considerable extent. A challenge from general disagreement along these lines that was not susceptible to this dual response would have to be one in which there was strong evidence that the subjects in question were thinking clearly, sufficiently intelligent, and relevantly informed, and *also* that, when they uttered sentences like ‘in this story, Joe knows that Mr. Pumbleton isn’t expecting him yet,’ their word ‘knowledge’ means knowledge, and that they were engaging with the same story we were. This set of circumstances is not obviously possible; certainly, the experimentalists have not presented evidence that it actually obtains.

I conclude that the case for extreme pessimism is weak. The experimentalist challenges can be met.

Part III: Knowledge, Counterfactuals, and Imagination

> *A person will worship something, have no doubt about that. We may think our tribute is paid in secret in the dark recesses of our hearts, but it will out. That which dominates our imaginations and our thoughts will determine our lives, and our character. Therefore, it behooves us to be careful what we worship, for what we are worshipping we are becoming.*
>
> *—Ralph Waldo Emerson*

Chapter 6
Knowledge, Counterfactuals, and Imagination

1. Apparent Connections Between Knowledge and Counterfactuals

The project of this final chapter is to articulate and explore a connection between knowledge and counterfactuals. Ultimately, I will defend a contextualist account of each. The starting point for my investigation consists in three observations about apparent connections between the two domains.

1.1. There appear to be necessary counterfactual conditions on knowledge.

Although particular attempts to develop the idea have proven controversial, as they are susceptible to apparent counterexample, many epistemologists have been attracted to the idea that a certain counterfactual relation between a subject and a proposition is a necessary condition for knowledge, by that subject, of that proposition. Robert Nozick, for instance, proposed a *sensitivity* requirement on knowledge: S knows that *p* only if, were *p* not the case, S would not believe that *p*. As many critics have observed, this requirement appears seriously problematic—for instance, combined with anti-skepticism, it appears to entail the failure of single-premise closure. (I know that I see Laura—if I didn't, it wouldn't appear to me that I did; but I don't know that I don't see a perfect imposter duplicate Laura—if I did, I'd still think that I saw Laura.) Nevertheless, the pre-theoretic appeal of the sensitivity condition on knowledge must be explained. There is something cogent about the argument: *how can you know that p? You'd think that p even if p weren't the case!*

Alternative counterfactual necessary conditions on knowledge have also been proposed. Ernest Sosa (1999) suggests, for instance, a counterfactual-based conception of *safety* as a necessary condition for knowledge: S knows that *p* only if, were S to believe that *p*, *p* would be the case. (Other, non-counterfactual, conceptions of safety have also been proposed by various authors, including Sosa himself. See §5.4 below.)

In §5 below, I will offer an explanation for the appeal of such counterfactual conditions on knowledge. My explanation will be that, apparent counterexamples notwithstanding, both safety and sensitivity are, so understood in terms of counterfactuals, genuine necessary conditions for knowledge.

(This is not to commit to any sort of JTBX-style analysis of knowledge; I do not think that the correct theory of knowledge will include safety or sensitivity as conjuncts. The claim is only the literal one: knowledge entails safety and sensitivity; necessarily, all cases of knowledge are cases of safe and sensitive belief. See Williamson (2003, 2-5).)

1.2. There are isomorphic puzzles about knowledge and counterfactuals.

A second connection between knowledge and counterfactuals requires explanation. Each gives rise to a certain sort of puzzle; these puzzles have striking similarities. The knowledge puzzle is widely-discussed—it is a standard skeptical paradox. Non-skeptics want to admit (1):

(1) I know that Laura is the person here.

However, it is intuitive to accept (2):

(2) I have no way to know that the person here isn't a perfect Laura imposter.

Nevertheless, the conjunction of (1) with (2)'s denial is unappealing:

(3) I know that Laura is the person here, even though I have no way to know that the person here isn't a perfect Laura imposter.

The conjunction, (3), is implausible on its face—we may also argue against it directly by invoking single-premise closure: If I know that Laura is the person here, then I can know that the person here isn't a perfect Laura imposter by deducing this from the known fact that Laura is the person here.

These three claims together, (1), (2), and the negation of (3), appear straightforwardly inconsistent, but each individually enjoys a certain plausibility. The potential moves in response are familiar: one may embrace skepticism, denying (1), or embrace the further knowledge, denying (2), or deny closure, accepting (3). Or, the contextualist response is to posit a kind of equivocation, suggesting that the use of 'knows' in (1)-(3) is not univocal, and that (1), (2), and the negation of (3), properly understood, are true and consistent.

This knowledge puzzle has an analogous puzzle for counterfactuals. The Count insists that the closet be opened; the Countess reluctantly acquiesces, and out comes the maid. Both Count and Countess are surprised—the Count, because he jealously suspected his wife of hiding a man in the closet; the Countess, because she *did* hide a man in the closet. The Countess is privately glad that the man somehow swapped places with the maid. Why is she glad? Because, she truly says to herself:

(4) If the Count had found a man, he would have become violent.

We may stipulate that the Count's temperament is such that (4) is true. Nevertheless, it is not the mere fact of there being a man found in the closet that would have resulted in the Count's fury; other factors have important roles to play. For instance, if he had found a man who immediately shot him dead, the Count would not have had time to become violent. That is to say, (5) is true:

(5) If the Count had found a man, the man could have immediately shot him dead; if he had found a man who immediately shot him dead, then he would not have become violent.

The puzzle arises from juxtaposing our verdicts about (4) and (5). As in the knowledge case above, we can present the difficulty in two ways. For one thing, it just flatly sounds contradictory to conjoin (4) with (5):

(6) If the Count had found a man, the man could have immediately shot him dead; if he had found a man who immediately shot him dead, he would not have become violent; but, if he had found a man, he would have become

violent.

Something is surely wrong; (6) is unacceptable, just as (3) was; an obvious candidate explanation for this unacceptability is that is that it is a contradiction.

If we take on a bit more theoretical baggage, we may derive the difficulty more formally. If the Count had found a man, he (at least) *might* not have become violent—he might, after all, have been immediately shot dead. If we treat the might-counterfactual as the dual of the would-counterfactual, as Lewis does, then this is equivalent to the negation of (4):

$$(\text{Man} \Diamond\!\!\rightarrow {\sim}\text{Violent}) \Leftrightarrow {\sim}(\text{Man} \,\Box\!\!\rightarrow \text{Violent})$$

We have again four choices—each corresponding to a familiar move in the parallel puzzle about knowledge. First, we may reject the attractive (4), sticking to our intuitions about (5). Alan Hájek (manuscript) suggests this approach; the corresponding move for knowledge is to embrace radical skepticism. Alternatively, we may stick to (4) and swallow the unwholesome negation of (5). No one, so far as I know, takes this position for counterfactuals—it is extremely unappealing—but its parallel in epistemology, according to which we insist that we do, contrary to intuition, know various skeptical scenarios not to obtain, is popular. Or we may, like the knowledge contextualist, attempt to find a hidden, shifty path that allows us to retain all the natural judgments.

The structural parallels between these two puzzles is a second connection between knowledge and counterfactuals for which I mean to provide an explanation.

1.3. We come to know counterfactuals by inferring in imagination.

If I know that you struck the match, then I can infer, on this basis, that the match lit, and come to know that the match lit. We may say that there is a *good inference* from *you struck the match* to *the match lit*. A similar cognitive phenomenon occurs when I evaluate counterfactual states of affairs and come to realize that *if you had struck the match, it would have lit*. On this, compare Williamson (2007):

> [W]e have various propensities to form expectations about what happens next: for example, to project the trajectories of nearby moving bodies into the immediate future (otherwise we could not catch balls). Perhaps we simulate the initial movement of the rock in the absence of the bush, form an expectation as to where it goes next, feed the expected movement back into the simulation as seen by the observer, form a further simulation, and so on. If our expectations in such matters are approximately correct in a range of ordinary cases, such a process is cognitively worthwhile. The very natural laws and causal tendencies our expectations roughly track also help to determine which counterfactual conditionals really hold. Thus some reliability in the assessment of counterfactuals is achieved. (148-49)

I will have more to say about the nature of inferences in imagination in §4 below.

Providing explanations for these three connections between knowledge and counterfactuals comprise central desiderata for the project of this chapter. I will begin by defending a form of contextualism about knowledge, then suggest that similar considerations motivate a related contextualism about counterfactuals. I conclude with a discussion of how these twin contextualisms help to explain the connections highlighted above.

2. Contextualism About Knowledge

2.1. What is David Lewis's contextualism?

According to David Lewis's brand of contextualism, knowledge attributions are universal generalizations. Lewis (1999) gives us this account of knowledge:

> S knows proposition P iff P holds in every possibility left uneliminated by S's evidence. (421)

Evidence, in Lewis's sense, is mentalistic: sense data, apparent memories, beliefs, and so on. A subject's evidence *eliminates* a possibility if it entails that that possibility is false—so any possibility in which I don't have the evidence I actually have is one that my evidence eliminates.

This is a contextualist account of knowledge, because the ‘every’ quantifier, like English quantifiers generally, has a context-sensitive scope. (Who must be present in order to make true an utterance of ‘everyone is here’? Not everyone in the universe—only some certain subset of them. Who is in that subset is determined partially by my intentions and the conversational context.) So, on Lewis’s view about knowledge, there is a set of possibilities that grows and shrinks according to conversational context; our knowledge attributions are true when the subject’s evidence eliminates all of the members of that set in which the object of knowledge is false. As Lewis puts it,

> S knows that P iff P holds in every possibility left uneliminated by S’s evidence—Psst!—except for those possibilities that we are properly ignoring. (425)

This is not meant as a change from the principle formulated above—it only makes explicit the fact that this ‘every’ is a restricted quantifier.

This is only a framework for an account of knowledge; much more must be said about which possibilities must be eliminated, and about how conversational context affects the relevant set. Lewis gives four rules that are meant to identify which possibilities are the ones in the domain of the universal. The *Rule of Actuality* demands that the actual world is always in the domain; this entails the factivity of knowledge. The *Rule of Belief* has it that a subject’s belief world is in the domain; we never properly ignore that which the subject believes to obtain. The *Rule of Resemblance* has it that worlds sufficiently similar, in a salient way, to a world in the domain are also in the domain. This rule is meant to explain many Gettier cases: when I look at a real barn in fake barn country, the domain includes, by the Rule of Actuality, the case where I am looking at a barn-looking structure in fake barn country, and so, by the Rule of Resemblance, it also includes the case where I am looking at a fake barn. Since my evidence doesn't eliminate this case, I don't know that I'm looking at a barn, even if the possibility that I'm in fake barn country never occurs to me. Finally, the *Rule of Attention* has it that any cases we are considering as possibilities are in the domain. It is primarily this rule that gives Lewis's account its contextualist feature—for what cases are treated as possibilities depends party on the conversational context. Often, we treat cases as relevant possibilities only once they've been mentioned. (The rule of resemblance may also have contextualist implications, if salient similarity is context-sensitive.)

2.2. Lewis gives a theory of both knowledge and 'knowledge'.

Contextualism is a semantic thesis about the word 'knowledge'; according to contextualism, the semantic value of this word varies according to conversational context. That is to say, among the factors that influence the truth value of an utterance of 'Tamino knows that Papageno is unreliable' are not only facts about Tamino and Papageno, but also facts about the conversational context of the speaker.

There is a sense, then, in which it my sound like a mistake to understand contextualism as an account of knowledge; it's really a thesis about the English word 'knowledge'. So we may be tempted to think of Lewis's statement of contextualism as committing a category error, confusing the question whether S knows that *p* with the question of whether an utterance of the form 'S knows that *p*' is true. Of course it is true that S knows that *p* just in case 'S knows that *p*' is true—but these are different propositions; they have different subject matters and different modal properties, even though they are biconditionally related *a priori* in actuality. (This is parallel to the point made against Stich in my chapter 5, §4.1, p. 117.)

An analogy can motivate the worry. Contextualism about indexicals is correct; whether an utterance of 'Papagena is here' expresses a truth depends not only on

Papagena's location, but also on that of the speaker. But it would be a mistake to say that, for instance, whether Papagena is here depends on where I am. So likewise, the worry goes, it is a mistake to say that whether Tamino knows that Papageno is unreliable depends on what possibilities we're considering. Lewis's treatment, unfortunately, is not careful about this distinction.

However, the issue is complicated by the fact that Lewis's story about knowledge—and mine, which will be related to his—goes beyond the mere claim of contextualism. It is also an account of knowledge. What can this mean, to be an account *of knowledge*, since the claim of contextualism is that there is not one but many different relations picked out by the word 'knowledge'?

Let's back up a step. An invariantist thinks that the truth of 'Tamino knows that Papageno is unreliable' depends on a number of factors: certainly on Tamino's brain state and whether Papageno is reliable, and probably some sort of connection between the two. It's open to controversy which facts play roles here; Jason Stanley (2005), for instance, has argued for the controversial thesis that Tamino's practical interests have an important role to play here, so that whether the sentence is true depends in part on whether Tamino is relying for anything important on Papageno's unreliability. Internalists and externalists may disagree about whether hidden features of Tamino's environment are factors.

What distinguishes invariantists is that facts about the conversational context in which the sentence is uttered do not impact the truth value of knowledge claims. We may understand Lewis's account as claiming that 'knowledge' picks out a single predicate that includes as a hidden argument, the context-sensitive 'all'.

Consider an analogy: someone is *undefeated* just there's no contest he's lost. This 'no' is a restricted quantifier, whose extension depends on the conversational context; in ordinary contexts, it's true to say that 'Mohammed Ali was undefeated,' even though he sometimes lost at Scrabble. Contextualism about undefeatedness is correct, even though we can give an account of undefeatedness; we needn't ascend to the metasemantic level to state the view about the nature of undefeatedness.

So likewise with Lewis's contextualist account of knowledge, and the one I will go on to defend.

2.3. Lewis's account is almost correct.

I believe that Lewis's contextualism has much to commend it. However, I do think that at least one modification needs to be made. Lewis's criterion for knowledge that *p* is met any time my evidence entails that *p*, relative to the relevant set of possibilities. Notice, then, that Lewis's account entails that evidence is luminous—that any time a subject has evidence E, she is in a position to know that she has evidence E. In fact, Lewis's account entails the stronger claim that any time a subject has evidence E, she *does* know that she has evidence E. For suppose that S has E. There are no possibilities (in the relevant context set, or even beyond) in which she has the evidence she actually has, and she also fails to have E. This follows straightforwardly from the fact that, for any evidence E, that the subject has E entails that E. Therefore, our subject, having E, thereby knows she has E. This is implausible, both on its face, and in light of Timothy Williamson's (2000, 93-8) argument that no nontrivial states are luminous.

I therefore propose the following modification of Lewis:

> S knows that *p* just in case, for some evidence E, (i) S believes that *p* on the basis of E, and (ii) all the E cases are *p* cases (Psst!—etc.).

This is the view I will defend, although it will require considerable clarification below. My proposal is very much in Lewis's spirit, and retains the context-sensitivity in the quantifier in my condition (ii). It adds an invocation of basing—something it would be surprising to see the correct theory of knowledge do without.

2.4. My account avoids Jonathan Schaffer's argument against relevant alternative theories.

My account also avoids the 'Missed Clues' objection that Jonathan Schaffer (2001) has raised against Lewis, and against relevant alternatives accounts generally. Schaffer invites us to consider a person who doesn't know that canaries can be distinguished from goldfinches by their wing color. The subject sees what is in fact a goldfinch, and sees that it has black wings. But he cannot tell whether it is a canary, even though in fact, canaries do not have black wings.

Schaffer's objection to Lewis is that the subject's evidence eliminates the possibility that it is not an (ordinary) canary, and so his account wrongly implies that the subject knows that it is not an ordinary canary. This strikes me as correct, so far—Schaffer's objection to Lewis here is similar to mine above.

But Schaffer goes further—he claims that the objection generalizes to any relevant alternatives approach to knowledge. He divides relevant alternatives approaches to his problem into two categories: those that attempt to deny knowledge by including unusual black-winged canaries as relevant alternatives, and those that attempt to deny that the subject's evidence eliminates the normal canary possibility. Schaffer rejects the former approach as allowing as relevant far-fetched skeptical hypotheses, and the latter as absurdly implying that the subject cannot know that the bird has black wings.

My account of knowledge highlights an alternative that Schaffer does not consider. The mutant canary is irrelevant, and the ordinary canary is relevant and eliminated by evidence—but the subject still does not know he is faced with a goldfinch, because he fails to form the corresponding belief in response to the appropriate evidence. This does not result in the conclusion that he does not know the bird has black wings—presumably, he does believe that, based on his visual experience as of black wings.

My account embeds Lewis's as a merely necessary condition. My additional condition is not met in the case Schaffer presents, and so his objection does not generalize to my approach.

2.5. Jason Stanley's objection to quantifier-based contextualism is unsound.

Jason Stanley's (2005) recent book, *Knowledge and Practical Interests*, contains a powerful critique of various forms of contextualism about knowledge. I believe that Stanley's objections to many forms of contextualism—such as views that attempt to treat knowledge along the lines of gradable adjectives—are sound. But in this section, I argue that Stanley's arguments do not impugn the version of contextualism I am endorsing.

One of the appealing features of this kind of contextualism is that it can explain skeptical intuitions—when skeptical possibilities are raised, our quantifiers' domains broaden to include them, and so we can no longer assert the same knowledge-sentences we could in non-skeptical contexts. That is, much of contextualism's support comes from its purported ability to resolve puzzles like the one given in §1.2 above.

Stanley (2005, pp. 62-65) has an argument against this approach. Quantifiers, Stanley says, can shift domains in and out in the middle of conversational contexts; we are happy to accommodate speakers who change the scopes of their quantifiers midway through conversations or even sentences. So, for instance, there is nothing problematic with:

> Every sailor waved to every sailor. (60)

(In a context in which two ships of sailors have just passed one another.) We accommodate, treating each 'every' as having a different scope, and so not requiring the absurdity that every sailor waved to himself.

Stanley also gives us this dialogue to illustrate how easily quantifiers can shift midconversation:

> A. *Every* van Gogh painting is in the Dutch National Museum.
>
> B. That's a change; when I visited last year, I saw *every* van Gogh painting, and *some* of them were definitely missing. (65)

I've emphasized the quantifiers: the first and third range over all the van Gogh paintings in existence; the second ranges only over those that were in the museum last year.

Since quantifiers shift so readily, Stanley says, Lewis's view should predict that knowledge attributions should also shift very easily. So according to Stanley, the view should predict that there is nothing wrong with 'abominable conjunctions' like:

> S knows that he has hands but S does not know that he is not a handless brain in a vat.

If knowledge attributions are just quantifiers, Stanley says, then why don't we adjust the understood domain restrictions to make this unproblematic? The objection to contextualism, then, is twofold: first, it wrongly predicts that such abominable conjunctions express truths, and second, it fails to resolve the skeptical paradox.

But this is unfair to the contextualist. Stanley has proven that sometimes we accommodate assertions by shifting domains; that doesn't mean that we have unlimited flexibility in this respect. It's not the case that our domains shift to *whatever* makes our utterances true in all cases. And indeed, in the apparently-relevant cases, domain shifting does not render the natural reading true. Take Lewis's gloss on the abominable conjunction above:

> All of S's uneliminated possibilities are hand-possibilities, but some of S's uneliminated possibilities are handless-brain-in-a-vat possibilities.

This conjunction is just as abominable as the knowledge one. It's a mouthful, so look at a more cognitively accessible conjunction with the same form:

> All of the bottles are on the table, but some of the bottles are in the fridge.

Because these quantifier-conjunctions are abominable, Lewis's theory does not predict that the relevant knowledge-conjunctions are unabominable. It is no embarrassment, then, that the knowledge-quantifiers don't shift in a way to make abominable conjunctions true.

2.6. The contextualist treatment resolves the skeptical paradox.

This contextualist treatment of knowledge resolves a familiar form of a skeptical paradox:

(1) I know that I have hands.

(2) I cannot easily know that I'm not a handless brain in a vat.

(3) If I know that I have hands, I can easily know that I'm not a handless brain in a vat.

Each premise is plausible, but their conjunction appears contradictory. The contextualist account allows each premise consistently to be true; an utterance of (1) expresses a truth when made in normal, non-skeptical contexts, because cases in which the subject is massively deceived are not part of the domain. (There are cases in which he has no hands, but these are cases his evidence eliminates.) An utterance of (2), however, brings attention to the skeptical hypothesis; by the Rule of Attention, then, the domain expands to include that hypothesis, and the knowledge claim is false: there is a case in the domain, uneliminated by the subject's evidence, where he has no hands. Likewise for (3); the conditional is true when evaluated at any given context. (The natural context for (3) is probably the skeptical context.)

So the paradox is diffused by assimilating it to this non-paradox:

(1') I drank all the beer.

(2') I didn't drink the beer we left in the store.

(3') If I drank all the beer, then I drank the beer we left in the store.

Note also that this treatment explains some dynamic features of knowledge attributions; to claim (1) or (1') is, in many contexts, felicitous, and (2) and (2') may well follow in the conversation. But statements of (1) or (1') are much more problematic when given *after* (2) and (2'). This apparent dynamic feature is evidence that there is a contextualist element at work in the case of knowledge.

2.7. My account is independent from problematic features of Lewis's account.

Some of the particulars of Lewis's view have proven controversial. Happily, many of these particulars are inessential to the broader project. In this section, I highlight four ways in which we may generalize from Lewis's framework.

2.7.1. Lewis's Rules

My proposal so far is a modification of Lewis's; it offers an account of knowledge in terms of a context-sensitive 'all cases'. It is uncontroversial that the semantic value of 'Tamino's evidence eliminates all the cases where Papageno is reliable' depends in part on its conversational context. English quantifiers like 'all' are context-sensitive.

Lewis goes to some effort to articulate how the conversational context, and the circumstances of the subject, determine the relevant domain. He posits a system of rules, alluded to above: the rules of actuality, attention, belief, and resemblance place constraints on which possibilities must be included; the rules of reliability, method, and conservatism articulate (defeasible) conditions that may be excluded. These particular rules have proven controversial.

For instance, Stanley (2005) complains that, in order to avoid counterintuitive consequences, the modal force of the rules of actuality and belief must be different from that of the rule of attention (112-13).

And indeed, some of the consequences Lewis claims for himself strike many philosophers—including myself—as implausible. He says that the mere mention of a skeptical possibility thereby *forces* it into the domain; that the mere act of considering possibilities destroys knowledge. (Hence, 'elusive'.) He is forced to describe his own project, then, as an attempt at 'saying what cannot be said' (1999, 566).

The framework against which I suggest we consider all of these challenges is this. Lewis offers, in 'Elusive Knowledge', two separate views: (a), an account of knowledge, in terms of the context-sensitive 'all cases', and (b), an account of the context-sensitivity of the latter, for which he posits a system of rules. In my view, Lewis's (a) represents a well-motivated insight, and something in the ballpark of it is true; by contrast, his (b) represents an extremely ambitious further project whose execution may be deeply flawed. Although I think that Lewis's rules gesture at plausible features of the dynamics of conversation, it is doubtful that any system so simple could be better than very approximately correct.

Consider a parallel in an evaluation of the sentence, 'Papageno finished the wine.' Plausibly, a context-sensitive quantifier may be unpacked from this sentence: it is true just in case, roughly, Papageno drank some wine such that, after he drank that wine, *all* of the wine had been drunk. I've here given an account of finishing that is parallel to the account of knowledge Lewis gives in part (a) above. It's context-sensitive, because the question of whether some particular bottle of wine (say, one that Papageno didn't know about) counts as part of *all* of the wine depends on the conversational context of utterance.

If we felt ambitious, we could try to fill in part (b) of the story as well, articulating a theory of how the circumstances of evaluation, and the facts about Papageno, determine which wine counts for the purpose of whether he drank 'all' the wine. Rules corresponding to Lewis's rules might appear attractive: any wine that we're thinking about counts; any wine that is obviously available to Papageno counts; any wine whose availability is relevantly similar to wine that is obvious to Papageno counts, etc.

Any list of rules that captures even most of the data will be rather baroque and invite charges of ad hocery. Some of the rules will cite features of the speaker (which wine is salient in the conversation), while some will cite facts about Papageno of which the speaker might be ignorant (which wine Papageno intends to drink tonight). This shows that the correlate of Lewis's project (b) is complicated and difficult to formalize in terms of simple rules—not that the correlate of (a) is misguided.

2.7.2. Attending and Ignoring

Lewis's treatment, according to which a possibility is forced into the domain just by being mentioned, which I suggested above to be implausible, has a parallel possible view about 'all the wine': any wine that has been mentioned or thought about in the context of the speaker is forced into the relevant domain, even if the speaker doesn't want it there, and it could otherwise have been legitimately ignored. Consider a case. Papageno has been enjoying a feast, which included two bottles of wine; he's just drunk the last of the second bottle. It is natural for someone to describe the case thus: 'Papageno finished the wine.' All of the wine in the domain is now gone.

But suppose that our speaker has an annoying conversational partner, who insists on bringing up irrelevant wines. He talks about Sarathustra's private stock of wine, which Papageno knows nothing about, and to which he has no access. According to a naive formulation of something parallel to Lewis's Rule of Attention (*i.e.*, any wine to which the speaker is attending is forced into the domain), we might have to say now that this wine is part of domain, and that it is not true to say, 'Papageno finished the wine.'

This result is implausible. The mere mention of some wine does not force it into the domain—or at least, it does not do so in a way that we cannot cancel. A perfectly legitimate response to the annoying conversational partner who brings up Sarathustra's wine runs: 'Yes, I know that Sarathustra has more wine. But I'm not talking about that wine. I'm talking about the wine that is part of the feast. Papageno finished the wine.' In other cases, cancellation is not even necessary. Suppose that you and I are at dinner, discussing Papageno (who is far away from us). We see that his glass is empty; you suggest, 'maybe Papageno will pour another glass now.' I reply, 'no, Papageno has already finished the wine.' Papageno's last bottle is empty, so I speak truly even if, as we speak, I am opening a new bottle at our table and pouring you a fresh glass. Our wine—to which we are attending—is outside this domain.

Insofar as this sort of move is plausible, Lewis's Rule of Attention may need to be weakened. In fact, I think that this is so, but I will not argue it further. The point I wish to emphasize is only this: the particular machinery that explains how the context-sensitive quantifier's domain depends on context will be complicated, and there is room, consistent with the account of knowledge I defend, to disagree on these sorts of particulars.

It is also worth remembering that the particular challenges of articulating which cases count as part of 'all' the not-*p* cases are general challenges; it's also difficult to articulate which wine counts as part of 'all' the wine.

Absent clear statements of the rules governing the relationship between conversational context and domain—a compelling answer to project (b)—do we sufficiently understand the content of the answer I defend to project (a)? Hopefully so. We manage to follow much language for which we're unable to articulate the relevant rules.

2.7.3. Evidence

Another possible generalization of Lewis ought to be emphasized. Lewis's account of knowledge invoked evidence; my modification of Lewis does too. It is controversial which propositions are part of one's evidence; Lewis clarifies that his evidence is internal: perceptual experience, apparent memory, and the like. (He is sanguine as to just what belongs on the list—'If you want to include other alleged forms of basic evidence, such as the evidence of our extrasensory faculties, or an innate disposition to believe in God, be my guest'—but it seems clear that he thinks, whatever counts as evidence, it must be at least introspectively recognizable.) (553)

In defending a cousin of Lewis's view, I don't mean to commit to his internalist conception of evidence. There are two lines of defense available—one mundane, and one exciting. My strategy is to embrace the exciting, and to fall back, if necessary, on the mundane.

The mundane defense is to understand 'evidence' in this account of knowledge as a term of art. Suppose that Timothy Williamson's (2000, 203-7; 2007, 206-15) externalist account of evidence is correct, and that everything I know counts as part of my evidence; then, in stating the view about knowledge, we should invoke not all our evidence, but all our evidence*—evidence* is what Lewis thought evidence was. We could, and probably should, write out the misleading word, thus:

> S knows that *p* just in case, for some set of perceptual and/or memory experiences {e_1, …, e_n}, (i) S believes that *p* on the basis of {e_1, …, e_n}, and (ii) all the cases where {e_1, …, e_n} obtain are *p* cases.

This is the mundane way to render the account of knowledge I propose compatible with externalist theories of evidence. The exciting way to render the view about knowledge compatible with externalism about evidence is to keep the original formulation in terms of evidence, and allow the relevant 'evidence' to be characterized by the correct theory of evidence, whatever that turns out to be.

We might worry about circularity or triviality results if we are open to certain accounts of evidence. Consider Williamson's again—your evidence is all and only that which you know. Can we plausibly plug even this knowledge-based account of evidence into this account of knowledge? In a word: yes. Two observations mitigate against the counterintuitiveness of this response. First, I am not intending to offer an *analysis* of the concept knowledge, breaking it into components that are conceptually prior to knowledge. This project strikes me as both hopeless and uninteresting. Second, my introduction of a basing requirement in §2.3 makes the account much friendlier to a knowledge-first picture of evidence. Here again is Lewis's own original formulation:

> S knows that *p* just in case S's evidence eliminates all the not-*p* cases.

Given Williamson's account of evidence, Lewis's account becomes all but vacuous:

> S knows that *p* just in case S's knowledge entails that *p*.

(This isn't *quite* vacuous, because it rules out the case where you fail to know that *p* because you fail to recognize the truth that *p* is entailed by your knowledge; of course, this is the sort of knowledge that I objected to in §2.3.) However, the invocation of a basing requirement changes the shape of the view. On my account,

> S knows that *p* just in case, for some evidence E, (i) S believes that *p* on the basis of E, and (ii) all the E cases are *p* cases.

If we plug in E=K, we get:

> S knows that *p* just in case, for some known propositions $\{e_1, \ldots, e_n\}$, (i) S believes that *p* on the basis of $\{e_1, \ldots, e_n\}$, and (ii) all the cases where $\{e_1, \ldots, e_n\}$ are true are *p* cases.

This principle is not vacuous in the way the previous one was. It places a genuine constraint on knowledge: in order to know that *p*, you must believe that *p* on the basis of some sufficient evidence. This framework leaves open the possibility that there is basic knowledge; this would be the limiting case in which the evidence is known on the basis of itself.

On this treatment, evidence will be context-sensitive, too. This is no vicious circularity. My knowledge depends on whether that which I may legitimately rely upon is strong enough to eliminate all the relevant alternatives; the conversational context of the knowledge attribution affects not only which alternatives are relevant, but also that which I may legitimately rely upon.

In fact, I find this combination of views—contextualism about knowledge and E=K—attractive and plausible. But my official view here is neutral about the nature of evidence; the limited point of this section is that we are not forced into Lewis's treatment of evidence.

2.7.4. Possibilities

S knows that *p* only when his evidence eliminates all the not-*p* cases. What is a case? For Lewis, it's something like a possible world, or a set of possible worlds. An odd implication, then, is that the are no distinct impossible cases; it is therefore very easy to know the proposition that *Hesperus is Phosphorus*; this proposition is identical to the proposition that *water is H_2O*, and that *triangles have three sides*. This is a problem common to philosophical views that individuate contents coarsely. (Stalnaker 1984.) Lewis's (1999) response, which is a standard one, is:

> [T]he necessary proposition is known always and everywhere. Yet this known proposition may go unrecognised when presented in impenetrable linguistic disguise, say as the proposition that every even number is the sum of two primes. (522)

This course-grained treatment of propositions is also optional on the present view. We may well take Lewis's approach, or we may choose instead to invoke structured propositions, where some propositions are impossible, or impossible worlds. I am inclined towards the latter options—we can thereby preserve the intuition that Lois doesn't know that Clark is Superman, and that the possibility that Clark is not Superman is uneliminated by her evidence. But I mean to remain officially neutral on this point, too.

The point of the preceding sections has been to motivate that many of the controversial elements to Lewis's account of knowledge are inessential to the core contextualist insight, which I endorse.

3. Counterfactual Contextualism

3.1. What are counterfactual conditionals?

Let us turn our attention now to counterfactuals. As I use the term, a *counterfactual conditional* is a sentence of the form, *if A had been the case, then C would have been the case*. 'Counterfactual' is probably a misleading title for this class of conditionals, as some counterfactuals have true antecedents; nevertheless, I will use the term in the way indicated, following the tradition of Lewis, Stalnaker, and other philosophers. (See Edgington 1995, 239-40.)

Counterfactual conditionals are contrasted with indicative conditionals, which are of the form, *if A is the case, then C is the case*. A famous and helpful example from Ernest Adams, cited by Edgington (1995, 237) illustrates the distinction:

> Indicative: If Oswald didn't kill Kennedy, someone else did.
>
> Counterfactual: If Oswald hadn't killed Kennedy, someone else would have.

These two claims may well vary in truth value (they do if conspiracy theories are false); therefore, indicatives are somehow different from counterfactuals.

It is clear that no truth-functional account of counterfactuals can be correct; the truth value of a counterfactual is not fixed by the truth value of its consequents and antecedents. (Consider: *if it had rained, I would have gotten wet*; but not: *if it had rained,* *I would have exploded*. If I neither got wet nor exploded, then these counterfactuals are both of the form F □→ F, but one is true and the other is false.)

An interesting project, then, is to articulate the truth conditions for counterfactuals.

3.2. What's right—and wrong—with David Lewis's account?

According to David Lewis's influential (1973) account, counterfactuals are *variably strict conditionals*. A counterfactual is true just in case some material conditional holds in all of the relevant set of worlds. The relevant set, according to Lewis, is the set of worlds where the antecedent is true that are *most similar* to the actual world—or, more precisely, more similar to the actual world than any worlds where the antecedent is false. So, to evaluate *if A were the case, then C would be the case*, we examine the sphere of A-worlds that differ less than any non-A worlds from the actual world, and check whether C is true in all of those worlds.

This explains some logical features of counterfactuals. For instance, it explains why counterfactuals are not closed under strengthening of the antecedent, the way that the material conditional is:

$$A \supset C \Rightarrow (A \,\&\, B) \supset C$$

But not:

$$A \mathbin{\Box\!\!\rightarrow} C \Rightarrow (A \,\&\, B) \mathbin{\Box\!\!\rightarrow} C$$

A counterexample to the latter is: *if I struck the match, it would light*, but: *if I soaked the match in water and struck the match, it would not light*. Lewis's explanation is that, although both counterfactuals are strict conditionals, they are strict conditionals over different domains. The first counterfactual's domain includes the nearest worlds where I strike the match—these are worlds in which I strike the match in the normal way. The second counterfactual's domain involves a different, more distant set of worlds. This is an attractive feature of Lewis's account.

An early and influential objection to Lewis's view focused on Lewis's invocation of similarity. Kit Fine (1975) observed that sometimes, counterfactual possibilities would have had dramatic effects on the course of history. So, for instance, *had Nixon pressed the button (to launch nuclear missiles), there would have been a nuclear war.* This counterfactual is very plausibly true—that's why we're glad he didn't press it—but it appears that Lewis's account will return the verdict that it is false. Consider the world where he presses the button, but the button malfunctions, and history runs its course as per actuality. This world is more similar to the actual world than the Armageddon world; so Lewis's account would consider this one the relevant world for evaluation of the counterfactual.

Lewis's (1986) response is that his invocation of 'similarity' is not to be thought of as overall macro-level similarity, but something more complicated—he gives, in 'Time's Arrow,' a baroque system of weighted rules, outlining how violations of our actual laws, of various sizes, negatively affect similarity. (It is right, I think, to be reminded at this point of the baroque system of rules he later gives in his attempt to characterize which possibilities count for his account of knowledge. See §2.7.1 above.) As a preface to this project, however, Lewis admits that he is attempting to articulate only one notion of similarity that we make use of in evaluating counterfactuals; some counterfactuals call for different similarity relations.

Of particular interest for my project is the suggestion that among the factors that influence what kind of similarity relation is relevant are facts about the conversational context. (Lewis uses the unfortunate term 'vagueness' in a way which can only be understood to mean 'context-sensitivity'.)

Lewis (1986) writes:

> What is going on, I suggest, can best be explained as follows. (1) Counterfactuals are infected with vagueness, as everyone agrees. Different ways of (partly) resolving the vagueness are appropriate in different contexts. Remember the case of Caesar in Korea: had he been in command, would he have used the atom bomb? Or would he have used catapults? It is right to say either, though not to say both together. Each is true under a resolution of vagueness appropriate to some contexts. (2) We ordinarily resolve the vagueness of counterfactuals in such a way that counterfactual dependence is asymmetric … (3) Some special contexts favor a different resolution of vagueness, one under which the past depends counterfactually on the present and some back-tracking arguments are correct. If someone propounds a back-tracking argument, for instance, his cooperative partners in conversation will switch to a resolution that gives him a chance to be right. … But when the need for a special resolution of vagueness comes to an end, the standard resolution returns. (4) A counterfactual saying that the past would be different if the present were somehow different may come out true under the special resolution of its vagueness, but false under the standard resolution. If so, call it a *back-tracking counterfactual*. Taken out of context, it will not be clearly true or clearly false. Although we tend to favor the standard resolution, we also charitably tend to favor a resolution which gives the sentence under consideration a chance of truth. (34)

It is easy to underestimate the significance of Lewis's concession here. Lewis is endorsing a limited form of contextualism about counterfactuals. Which notion of similarity is relevantly salient may change according to the conversational context; therefore, whether a counterfactual sentence is true depends not only on the antecedent, the consequent, and the state of the world, but also upon the conversational context. Here are some examples.

Lewis gives us the example of Caesar counterfactuals: 'if Caesar had been in command in Korea, he would have used the atom bomb.' This expresses a truth in some contexts. But in other contexts, we'll say instead: 'if Caesar had been in command in Korea, he would have used catapults.' Which context we are in depends on features of what we intend, and on what we've been talking about. (Only a very unusual context could license both sentences.)

When faced with a counterfactual, we imagine the antecedent. But just what we are to imagine is underdetermined by the antecedent alone; *how* are we to imagine Caesar in command in Korea? Should we imagine him having replaced General MacArthur? Or should we imagine the Holy Roman Empire having expanded to the east end of Asia?

The same question underwrites questions about backtracking counterfactuals. Shy Stan says, ‘I’m terrified of public speaking. If I were the White House spokesman, I’d give the worst press conferences.’ Quick Kevin retorts: ‘You’d never be White House spokesman unless you got over your public speaking difficulties, Stan. If you were the White House spokesman, you’d give great press conferences (because you’d’ve gotten over your difficulties).’ There is no real disagreement here; each may well be expressing a truth—even though the truths take the form of apparent contradictions.

Indicative conditionals may be understood to work in much the same way—the indicative mood signals an alternate way of filling out the scenario. (See McCawley, 1997, 90-91.) Indeed, in some circumstances, counterfactuals can even function in much the same way that indicative conditionals do. Here, I borrow from Dorothy Edgington (manuscript): You and I team up on a treasure hunt, and the organizer has given me a hint: ‘it’s either in the attic or the kitchen.’ We split up; I search the attic, and you search the kitchen. I find the treasure in the attic, then you ask me why I sent you to the kitchen. I reply: ‘Because if it hadn’t been in the attic, it would’ve been in the kitchen.’ I here speak truly, even if the organizer’s second choice for a hiding place was the basement.

Once we are convinced that this sort of limited contextualism is true about counterfactuals, it is worth investigating whether such context-sensitivity runs much deeper—and what sorts of general principles might underwrite all counterfactuals.

3.3. There is a strong case for contextualism about counterfactuals.

I pointed out in §1.2 above that the skeptical paradoxes about knowledge have analogues for counterfactuals. These counterfactuals seem true, but their conjunction seems abominable:

> (4) If the Count had found a man, he would have become violent.
>
> (5) If the Count had found a man, the man could have immediately shot him dead; if he had found a man who immediately shot him dead, then he would not have become violent.
>
> (6) If the Count had found a man, the man could have immediately shot him dead; if he had found a man who immediately shot him dead, he would not have become violent; but, if he had found a man, he would have become violent.

The knowledge contextualist resolves his puzzle by suggesting that the standards for knowing shift between the premises; the counterfactual contextualist, likewise, admits that (4) and (5) each express a truth, but that changing features of the conversational context prevent the apparent entailment from (4) and (5) to (6).

One bit of support for this approach is that pairs of counterfactuals like these appear to be sensitive to dynamic conversational features. Pairs of counterfactuals like (4) and (5) *can* be acceptably conjoined in some conversational contexts, called Sobel sequences. Suppose that Susanna relates her story to Figaro, saying:

> 'If the Count had found a man, he would have become violent. But of course, the man could have shot him dead; if he had found a man and died instantly, he would not have become violent (because he would have been dead).'

This bit of speech is unproblematic. (Perhaps one might worry about pragmatic considerations—why would she even mention such an odd possibility? But it is not difficult to cook up unusual contexts in which the utterance makes sense.) Things change dramatically, however, for *Reverse* Sobel sequences—suppose Susanna puts her counterfactuals into the opposite order:

> 'If the Count had found a man, he could have died instantly; had he found a man and died instantly, he would not have become violent (because he would have been dead). But of course, if he had found a man, he would have become violent.'

Something is deeply wrong with this utterance—something that goes beyond mere pragmatic oddity. It's not that we don't know why she's bothering to mention what she mentions—it's that we have difficulty understanding her at all. Susanna here seems flatly to be contradicting herself. This asymmetry has convinced some philosophers and linguists that there must be a dynamic element to counterfactuals. (See, for example, Gillies 2007; McCawley 1997.)

The same pattern emerges for knowledge attributions: 'I know I have hands; but of course, I don't know that I'm not a brain in a vat.' This conjunction is greatly preferable to the reversed version: 'I don't know I'm not a brain in a vat, but of course, I know I have hands.' This feature provides part of the motivation for my Lewisean contextualism about knowledge; to say I know I have hands is to say that all the (relevant) no-hand possibilities are eliminated. Once we're thinking about brains in vats, situations in which I only apparently have hands become relevant, and the knowledge attribution is no longer apt.

I suggested in §2.7.2 that it may well be easier to retreat to nonskeptical contexts than Lewis supposed. One way to do this might be to explicitly rule out the possibility: 'If I were a brain in a vat, I wouldn't have hands. *But that's silly; of course I'm not a brain in a vat.* I know I have hands.' If this sort of move is legitimate here, it may also be in the case of the counterfactual: 'If the Count had found a man, he could have been instantly killed. *But of course that's silly; Cheribino wouldn't have killed the Count.* If the Count had found a man, he would have become violent.'

This parallel suggests an account of counterfactuals:

> *If A were the case, C would be the case* is true just in case all of the A possibilities are C possibilities.

This 'all', like the one in my account of knowledge, takes a context-sensitive scope. So, we might say:

> *If A were the case, C would be the case* is true just in case all of the A possibilities are C possibilities (Psst!—except those possibilities we're properly ignoring).

This account is an example of the type laid out by McCawley (1997, 91).

Although Lewis (1973) provided the groundwork for a strong contextualist solution to the knowledge puzzle, he did not seem to think much of the corresponding solution to the counterfactuals puzzle, opting instead (1986) to further complicate his static similarity relation in an attempt to capture our ordinary judgments (even while admitting that he was attempting to capture only the 'standard resolution' for counterfactuals). The Lewis of 1973 called a strict conditionals view of counterfactuals in which the relevant class of worlds is contextually determined, 'defeatist', and accused such an approach as 'consign[ing] to the wastebasket of contextually resolved vagueness something much more amenable to systematic analysis than most of the rest of the mess in that wastebasket.' (13) It's not clear whether the Lewis of 1996 retained his contempt for attribution of context-sensitivity; was Lewis really willing to throw *knowledge* into that 'wastebasket', when counterfactuals were too good for it? Lewis's own reservations notwithstanding, a contextually variable strict conditionals analysis of counterfactuals is every bit as plausible as a contextually variable infallibilist analysis of knowledge.

The knowledge account was incomplete without a story about which possibilities were properly ignored; so likewise is this account of counterfactuals. I suggest that just the same possibilities are relevant in each case; the rules Lewis suggested for relevance with respect to knowledge—the rules of actuality, belief, resemblance, and attention—gesture approximately at the relevant worlds for the evaluation of counterfactuals. (As I suggested above, the rules themselves are almost surely imperfect, for both cases.) Put differently, knowledge attributions and counterfactual statements are different generalizations over the same sets of possibilities: when I say I know that *p*, I say that my evidence eliminates all the not-*p* cases in the set. When I say that if A were the case, C would be the case, I say that the material conditional, A ⊃ C, holds in all the cases in the set.

On this account, it is not difficult to see how to resolve our puzzle. When Susanna utters (4), she says that all (and here she restricts her quantifier, ruling out irrelevant cases as properly to be ignored) possibilities in which the Count finds a man are cases where the Count becomes violent. But once the possibility of the Count finding a man and immediately being killed is raised, it is no longer properly ignored. This because, at least, it is no longer ignored at all. Since this is a possibility in which the Count doesn't become violent, (4) is false as uttered. This explains why (5) is true, and why (4), which looked so good on its own, is unacceptable, once (5) has been asserted, and the Count-dying possibility is put on the table.

3.4. The same set supplies domains for knowledge and counterfactuals.

Why think that knowledge attributions and counterfactual utterances express universal quantifiers over the very same domains? The strongest case in favor of this identity comes in the form of particular cases: consideration of cases that are counterexamples to a knowledge generalization, when those cases are also counterexamples to the counterfactual generalization, undermine the knowledge attribution and the counterfactual utterance alike. Take, for instance, an ordinary context, in which both of these utterances express truths:

> 'Nancy knows that Cheney will continue to be Vice-President tomorrow.'
>
> 'If Bush died tonight, Cheney would be President tomorrow.'

The domain, in this nonskeptical context, does not include cases in which Cheney dies tonight. (These cases, we may suppose, are very distant and legitimately ignored.) But we may explicitly bring those cases into the domain; doing so undermines both sentences:

> 'Of course, Cheney could die suddenly tonight, but Nancy knows that Cheney will continue to be Vice-President tomorrow.'
>
> 'Of course, Cheney could die suddenly tonight, but if Bush died tonight, Cheney would be President tomorrow.'

Both of these sentences are false in their contexts—each is undermined in just the same way, by the consideration of a particular counterexample: a case where Cheney dies suddenly tonight. It is a counterexample to the knowledge attribution because it's a case where Nancy's evidence is as it actually is, but Cheney will not be Vice-President tomorrow; it is a counterexample to the counterfactual utterance because it's a case where Bush dies tonight, but Cheney will not be Vice-President tomorrow.

I am unaware of any data suggesting that the two domains should come apart; consideration of particular cases like this one provides some reason for thinking that they will shift together; general theoretical simplicity provides a second reason. A third positive reason to treat the relevant domains as coextensive is that doing so permits the attractive explanations for the phenomena I laid out in the introduction.

I turn now to these explanations.

4. Inferences, Knowledge, and Counterfactuals

I promised above to explore the theoretical similarities between inferences leading, in beliefs, to knowledge, and inferences leading, in imagination, to counterfactuals. Now is the time for this exploration.

4.1. When imagining, we often infer.

If I know that you struck the match, then I can infer, on this basis, that the match lit, and come to know that the match lit. (So, we are assuming that there are no relevant possibilities in which you strike the match and it does not lit; this is the case in ordinary circumstances.) We may say that there is a *good inference* from *you struck the match* to *the match lit*. A similar cognitive phenomenon occurs when I evaluate counterfactual states of affairs and come to realize that *if you had struck the match, it would have lit*. This similarity is explained by the accounts of knowledge and counterfactuals given above. What it is for there to be a good, knowledge-preserving inference from your striking the match to its lighting is for there to be no possibilities in which you strike the match and it does not light. Assuming that we hold the relevant context fixed, this is exactly what's required of the counterfactual: for all the striking possibilities to be lighting possibilities.

It is helpful to think of the inference from your striking the match to its lighting as more general than one that takes us from belief to belief; in evaluating the counterfactual, we *imagine* the antecedent, then reason according to the very same inference we use in the case where we believe it. For any given epistemic inference governing belief, there is an isomorphic cognitive inference governing imaginings. Call these inferences *hypothetical inferences.* Hypothetical inferences, involving imaginings, can be thought of, along with their corresponding epistemic inferences, involving beliefs, as special cases of a general pattern of *cognitive inference*. (I believe, but won't here argue, that the same is true of *practical inferences*, involving relations between desires.) The normativity of epistemic inference is based in the norms of cognitive inference. I will have only a little to say about those norms themselves, although we clearly have at least a good tacit understanding of them, which is why we are able to make epistemic assessments; I am interested in abstracting away from discussion of epistemic inference to cognitive inference in general.

Strict entailment is too strong a requirement for goodness of cognitive inference. I suggest that instead, we consider entailment *relative to the context set*, where the context set is characterized as outlined above.

It is helpful to revisit a few interesting features about the operation of the imagination. Psychological and philosophical investigation of the imagination in recent decades seems to have reached the following limited near-consensus about imagination: imaginative propositional attitudes are interestingly and importantly belief-like, but nevertheless comprise a distinct cognitive attitude from belief. Beliefs and imaginings can take the same kinds of contents. It is generally accepted among theorists about imagination that moves very much like the inferences that occur in belief also occur in imagination. Shaun Nichols (2006) summarizes the suggestion nicely:

> [W]hen people engage in imaginative activities, they often follow orderly inferential chains. When I read that Wilbur is a pig, I infer (in imagination) that Wilbur is a mammal. When I hear that Hamlet is a prince, I infer (in imagination) that he is not a member of the hoi polloi. These inferences track the kinds of inferences that I would have if I really believed that Wilbur was a pig and Hamlet was a prince. Such orderly inferences emerge on the scene very early in childhood. … [In a 1994 study by Alan Leslie, two-year-old children] are apparently drawing inferences over the contents of what they are pretending, and the inferences parallel the inferences that the children would have if they had the corresponding beliefs. (7-8)

To adapt from one of Leslie's cases: I imagine that the cup is full of tea, and I see that the cup is overturned above the bear, and I go on to imagine that the bear is wet—or if I don't, then I 'change my mind' and imagine that the cup wasn't full after all. This new imagining is the result of an inference—just the same cognitive inference I would perform if I believed that the cup contained tea, instead of merely imagining it. So imaginings, too, can be susceptible to inference. Imagining one proposition may cause me—in an inference-like way—to imagine another. And the structure of those inference-like imaginings is isomorphic to that of inferences in belief.

One might object: there is an important disanalogy between belief and imagination. Rationality demands proper inferences in the former cases, but there is nothing wrong with the person who imagines according to unusual patterns. It does seem as if there is a normative difference here. The person who believes that *p* and who knows full well that *p* comes along with *q* is irrational if he does not come to believe *q* as well. But the person who imagines *p* without imagining *q* does not seem thereby to be rationally suspect. I grant this disanalogy but dismiss it as unimportant. All-things-considered rationality is not essential to proper cognitive inference. Hypothetical inferences are not mandatory the way that theoretical and practical inferences are, but they can still be evaluated as instances of good or bad cognitive inferences (with the understanding that, depending on what you're trying to do, it may not be inappropriate to run a bad cognitive inference in imagination, or to fail to run a good one). But there is still a place for rational normativity in hypothetical inference.

4.2. This approach resolves Goodman's 'problem of law'.

This framework helps us to answer an old challenge in the philosophical literature about counterfactuals: Nelson Goodman's 'problem of law'. Counterfactuals demand a certain kind of connection that must hold between the antecedent (plus relevant cotenable conditions) and the consequent; the problem of law is the challenge to spell out what that connection is. Sometimes, we can *deduce* the conclusion from the antecedent; my competence with English and logic allows me easily to see that *if I were smaller than a breadbox, a breadbox would be larger than I*, or that *if I weren't a bachelor, I would be either non-male or married*. In these cases, the semantic facts and the laws of logic provide the connection between antecedent and conclusion. But this is not the case with many of our counterfactuals. Goodman (1983) writes:

> [E]ven after the particular relevant conditions are specified, the connection obtaining [between antecedent and conclusion] will not ordinarily be a logical one. The principle that permits the inference of

> That match lights
>
> from
>
> > That match is scratched. That match is dry enough. Enough oxygen is present. Etc.
>
> is not a law of logic but what we call a natural or physical or causal law. The [problem of law] concerns the definition of such laws. (13)

Goodman is impressed by the fact that for many counterfactuals, the truth of the counterfactual seems to be underwritten by the truth of the lawlike generalization of the counterfactual; we think that *any* time a match is scratched, so long as it is dry, in the presence of oxygen, etc., it will light. Goodman's contrasting case involves the change in his pocket: all the change in his pocket at *t* was silver, but the corresponding counterfactual is not supported: it's not the case that if this penny were in his pocket at *t*, it would have been silver. This is because the generalization, though true, is not a law. Goodman's challenge for counterfactual theorists, then, is once again to come up with a non-circular way to specify the relevant property: this time, lawfulness. Goodman also writes:

> When we say
>
> > If that match had been scratched, it would have lighted,
>
> we mean that conditions are such—i.e. the match is well made, is dry enough, oxygen enough is present, etc.—that "That match lights" can be inferred from "That match is scratched." (8)

The preceding discussion provides a way to understand 'inference' here without recourse to laws. There is a good cognitive inference from the match striking to the match lighting. That's why, if we do not change the background conditions, one who believes the former must believe the latter (or reject the former). The cognitive inference that we perform in reasoning with beliefs also has an important realization in hypothetical reasoning: it is how we come to know the counterfactual.

(Goodman may here be offering a semantics for counterfactuals: what it is to assert the counterfactual *is* to claim that the conditions permit the inference. If he is making such a claim, I wish to dissociate from it; I am claiming only that there is an important connection between the counterfactual and the inference: in particular, performing the inference is the characteristic way to come to know the counterfactual.)

Good cognitive inference, then, can take the place of Goodman's lawlike deduction. If I infer *q* from *p*, where that is an instance of a good cognitive inference, I can on this basis know that *p* counterfactually implies *q*. We know which inferences are appropriate, because they derive from cognitive inferences, the appropriateness of which we understand. If you know that the cup is full and that it is overturned over the bear, you can know that the bear gets wet. And in general, for any appropriate knowledge-preserving inference in belief, the corresponding inference in imagination is appropriate in the sense we seek. That is, just in case it is appropriate to infer the belief that C from the belief that A (relevant to some background conditions), it is appropriate (in this counterfactual-relevant sense) to infer the imagining that C from the imagining that A (along with the same background conditions).

This suggestion explains the points of agreement between Goodman's view and my own—when you infer by deducing from known laws, your cognitive inference is good—and also generalizes to cases that proved problematic for Goodman. (See Edgington 1995, 249-50.) If I know that I've just thrown the glass of wine in Amanda's face, but cannot yet perceive her reaction, I can appropriately infer—and thereby know—that she is angry. Suppose that I throw the wine in her face and am immediately pulled aside by a protective friend, so that I cannot see her. In that situation, I can know that she is angry by inferring it from the fact that I threw wine in her face, along with the relevant background conditions. It is just that inference that underwrites my knowledge of the counterfactual, *if I were to throw this glass of wine in Amanda's face, she would be angry*.

We may pithily approximate the point thus: we can know counterfactuals using inferences that would give us knowledge of the conclusion, given knowledge of the antecedent. We may understand knowledge of counterfactuals in terms of counterfactual knowledge. But counterfactual knowledge (knowledge in counterfactual situations) is useful here only as a heuristic for recognizing and thinking about good inferences. Good inference, not counterfactual knowledge, is the key to good counterfactual reasoning.

The advantages of emphasizing inference over counterfactual knowledge come out particularly when considering counterfactuals with possible but unknowable antecedents, like *if I had never been born, my sister would have been an only child*. The inference is good, even though I could never know the antecedent. On a counterfactual knowledge emphasizing view, we would be forced to evaluate the counterfactual, *if I knew I was never born, I could appropriately infer that my (actual) sister is an only child*. It's not obvious how, or whether, counterfactuals with impossible antecedents should be evaluated. At any rate, a view is preferable if it doesn't require us to commit to any particular evaluation of this very strange counterfactual to understand the perfectly ordinary one we began with. Furthermore, I avoid worries about circularity by citing good inference instead of counterfactual knowledge.

5. Safety, Sensitivity, and Knowledge

5.1. Sensitivity is an attractive necessary condition for knowledge.

As mentioned in the introduction to this chapter, an attractive necessary condition for knowledge is *sensitivity*:

Sensitivity: A belief that *p* on the basis of E is sensitive just in case, if *p* hadn't

> been the case, the subject wouldn't have believed that *p* of the basis of E.

This is my gloss of sensitivity, anyway; sensitivity is often presented without the invocation of the subject's basis for belief: S's belief that *p* is sensitive just in case, were *p* not the case, S wouldn't believe that *p*. I invoke the basis with an eye toward cases in which a subject believes that *p* on good grounds, but would believe that *p* on bad grounds, were *p* not the case. Here are two examples:

> (a) A wife believes, on the basis of excellent evidence, the truth that her husband is faithful. But she is psychologically unable to face difficult truths. So, had her husband been unfaithful, she would have self-deceived herself into irrationally believing him to be faithful.
>
> (b) The priest has gone to see whether the god is on the mountain; he climbs the mountain, sees the god, and comes to believe that the god is on the mountain. The god, anxious for his subjects to believe him to be on the mountain, plants a trap whenever he goes away, which will zap any visitors' brains. Their memories of having been to the mountain are erased, and they are given an unreflective belief that the god is on the mountain. So, had the god not been on the mountain, the priest would have had his memory wiped and left with the unreflective belief that the god is on the mountain.

The properly-formed beliefs of the wife and the priest, intuitively knowledge, are sensitive in my sense.

5.2. Arguments against sensitivity rely upon an invariantist semantics.

A sensitivity requirement occupies a central role in Dretske's and Nozick's treatments of knowledge. But sensitivity is often thought to be problematic as a requirement for knowledge. As I pointed out in §1.1 above, sensitivity, combined with anti-skepticism, appears to imply the failure of single-premise closure. And Ernest Sosa, among others, has provided, independent of this worry, apparent counterexamples to a sensitivity requirement for knowledge. Here is an influential one:

> On my way to the elevator I release a trash bag down the chute from my high rise condo. Presumably I know my bag will soon be in the basement. But what if, having been released, it still (incredibly) were not to arrive there? That

> presumably would be because it had been snagged somehow in the chute on the way down (an incredibly rare occurrence), or some such happenstance. But none such could affect my predictive belief as I release it, so I would still predict that the bag would soon arrive in the basement. My belief seems not to be sensitive, therefore, but constitutes knowledge anyhow, and can correctly be said to do so. (Sosa, 1999, 145-6)

The garbage chute case appears to be a case of insensitive knowledge. In this case, the closest worlds in which the bag won't make it to the basement are worlds in which the subject falsely believes, on the basis of his actual evidence, that the bag will make it to the basement. So, Sosa has provided a counterexample to this principle:

> **Need Nearest:** S knows that *p* on the basis of E only if, in the nearest possible worlds in which not-*p*, S does not believe that *p* on the basis of E.

Is the garbage chute case a counterexample to sensitivity as a requirement for knowledge? Recall that sensitivity invokes a counterfactual: S's belief that *p* on the basis of E is sensitive just in case, if *p* hadn't been the case, S wouldn't have believed that *p* of the basis of E.

A sensitivity requirement for knowledge is equivalent to **Need Nearest** just in case Lewis's semantics for counterfactuals is correct. Since Lewis's semantics is widely assumed to be correct in much of the relevant literature, it is not surprising that a counterexample to **Need Nearest** should often be assumed without argument to be a counterexample to a sensitivity requirement. But an alternative semantics—one we've seen to be in some ways preferable—is now on the table. How does sensitivity fare if we understand counterfactuals, and knowledge, as described above?

5.3. Sensitivity is unproblematic with context-sensitive counterfactuals.

I have placed the bag into the chute, and now know, on the basis of my experience, that it will shortly arrive in the basement. That is to say, my evidence rules out all the possibilities where the bag won't shortly arrive in the basement (except the ones we're properly ignoring). What of the counterfactual that would establish sensitivity? *If the bag were not to arrive shortly in the basement, I would not believe that the bag would arrive shortly in the basement*. This is true just in case all the possibilities in which the bag won't land in the basement (again, with a suitably restricted 'all') are cases in which I don't believe the bag will land in the basement. This is false; the relevant counterexamples are the ones Sosa discusses—cases where the bag, unbeknownst to me, gets stuck on the way down. So doesn't Sosa's counterexample generalize to sensitivity, even given my picture of counterfactuals?

Perhaps not. On my approach, there is a context-sensitive element to both knowledge attributions and counterfactuals—the evaluation of each depends on which possibilities we're properly ignoring. And the cases that falsify the counterfactual—the ones where the bag gets stuck in the chute—are exactly the ones we must properly ignore in order to go through with the knowledge attribution. The two claims that together appear to violate sensitivity, then, only hold if we switch contexts between them. When we consider the counterfactual, *what if the bag won't make it to the basement?*, we expand the class of relevant possibilities to include some in which the antecedent holds. Note that, once we've done this, we can no longer make the knowledge attribution; it is defective to say, 'if the bag weren't to make it to the basement, I'd falsely believe the bag would make it to the basement. But I know the bag will make it to the basement.'

This is a common pattern with context-sensitive discourse; here is an instructive parallel. Consider this extremely attractive principle:

EAP. All Φs are F only if all G Φs are F.

Someone sets out to provide a counterexample to this extremely attractive principle. He points to a fine jewelry store, filled with expensive diamonds. None of the merchandise is available for less than $1,000. 'Everything in this store is expensive,' he says, speaking truly. Then he adds, thinking of the wastebaskets, which contain such things as paper receipts: 'The *garbage* in this store is not expensive.' He speaks truly again. Then, conjoining his two true utterances, he claims to have found a counterexample to EAP—everything in the store is expensive, even though the garbage in the store is not expensive.

The fallacy is transparent. The domain of the quantifier 'everything' varies according to the conversational context. We may truly say that everything in the store is expensive only when we are restricting our quantifier to the merchandise—only when we are properly ignoring things like the contents of the wastebaskets; when we say that the garbage in the store is not expensive, we stop ignoring the garbage, so the domain broadens to include it. This is why we cannot give the two sentences in the opposite order: *The garbage in the store is not expensive, but everything in the store is expensive.* The imagined counterexample relies on illicit changes in conversational context. This is clearly the right response in this case—we do not give up the claim that everything in the store is expensive only if there are no groups of things that are inexpensive. That 'everything in the store is expensive' plausibly expresses a truth only when wastebasket contents are ignored; considering whether the wastebasket contents in the store are expensive demands us to consider the wastebasket contents—once we do, we admit that they're not expensive (and are no longer in a position to assert that everything in the store is expensive).

In just the same way, the garbage chute case needn't threaten a sensitivity requirement for knowledge; that the subject knows the garbage will be in the basement is only plausible when the garbage-getting-stuck cases are ignored; considering whether the relevant counterfactual (*if the garbage weren't to reach the basement, the subject wouldn't believe that it would*) is true demands us to consider the garbage-getting-stuck cases—once we do, we admit that the counterfactual is false (and are no longer in a position to assert that the subject knows that the garbage will reach the basement).

5.4. Should we prefer safety to sensitivity?

Sosa's argument, purportedly refuting a sensitivity requirement for knowledge, in fact refutes only the combination of a sensitivity requirement for knowledge and Lewis's semantics for counterfactuals. Since I have argued independently against the latter, Sosa's argument is not persuasive against a sensitivity requirement. This isn't yet to say that sensitivity *is* a requirement for knowledge—only that, for all Sosa has given us, there's no reason to take that possibility off the table. We may wonder whether to prefer a sensitivity requirement or an alternative, such as Sosa's *safety* requirement on knowledge. Here is Sosa's gloss on safety:

> Call a belief by S that *p* "safe" iff: S would believe that *p* only if it were so that *p*. (Alternatively, a belief by S that *p* is "safe" iff: S would not believe that *p* without it being the case that *p*; or, better, iff: as a matter of fact, though perhaps not as a matter of strict necessity, not easily would S believe that *p* without it being the case that *p*.) (Sosa, 1999, 146)

Let's add a basis requirement to each formulation, to match the one I added for sensitivity. Then we have, for Sosa's three formulations:

> S's belief that *p* on the basis of E is safe just in case:
>
> > S would believe that *p* on the basis of E only if it were so that *p*;
>
> or:
>
> > S would not believe that *p* on the basis of E without it being the case that *p*;
>
> or:
>
> > as a matter of fact, though perhaps not as a matter of strict necessity, not easily would S believe that *p* on the basis of E without it being the case that *p*.

The modification is to protect safety from the same sorts of cases mentioned above—

cases in which the subject responds properly to the evidence, but who could have behaved irrationally instead. Add to the two examples that the husband could easily have been unfaithful, or that the god could easily have been away, and we have failures of safety if we don't include a basis in the formulation, and cases of safety if we do.

In the garbage chute case, the relevant belief is safe because, although there are possible scenarios in which the subject believes as he does and does so falsely, none of those scenarios could *easily* obtain. The relevant worlds are too distant. Sosa thinks this to be a difference between safety and sensitivity, but I have argued that, in the cases where it is plausible that the subject knows that the garbage will make it to the basement, that belief is sensitive after all (although I admitted that, if we dwell on its sensitivity, the context will shift to one in which we must deny both the knowledge attribution and the claim to sensitivity).

In fact, we may draw a quite general relationship between sensitivity and safety. Sosa explains how he thinks they must differ:

> A belief is sensitive iff had it been false, S would not have held it, whereas a belief is *safe* iff S would not have held it without it being true. For short: S's belief B(p) is sensitive iff $\sim p \ \Box\!\!\rightarrow \sim\mathrm{B}(p)$, whereas S's belief is safe iff $\mathrm{B}(p) \ \Box\!\!\rightarrow p$. These are not equivalent, since subjunctive conditionals do not contrapose. (146)

And he elaborates with this footnote:

> If water now flowed from your kitchen faucet, it would *not* then be the case that water so flowed while your main valve was closed. But the contrapositive of this true conditional is clearly false. (152)

But, as before, my treatment of counterfactuals provides a debunking explanation of the apparent counterexample; when we consider the counterfactual Sosa gives, we consider a set of possibilities that doesn't include any cases where the water leaks through a closed valve; when we consider the contrapositive, *if water flowed from your faucet while the main valve was closed, water would not have flowed*, we turn to a new set of possibilities.

So the case is consistent with the principle that a subjunctive conditional is equivalent to its contraposition, when we hold the context fixed—it's just that sometimes, one formulation forces us into a context in which the other no longer expresses a truth.

Don't forget the helpful analogy with context-sensitive quantifiers: (not everyone has studied semantics but) everyone understands English; of course, some Chinese people don't understand English. This does not refute the duality of the universal and the existential. Or, to tighten the cases, go back to the water faucet case and work with the quantifiers I use to gloss the counterfactuals:

CF_1: 'If water were to leak, it would not leak from a closed faucet.'

Q_1: 'None of the possibilities where water leaks involve water leaking from a closed faucet.'

CF_2: 'If water were to leak from a closed faucet, water would not leak.'

Q_2: 'None of the possibilities where water leaks from a closed faucet involve water leaking.'

Each of Q_1 and $\sim Q_2$ can express a truth—Q_1 might be a useful thing to say if, for instance, one wanted to take steps to ensure that water wouldn't leak. Just make sure to close the faucet, and the water won't leak. We are ignoring cases in which the water manages to leak out of a closed faucet—our faucet is in good working condition, so these are reasonable cases to ignore. We stop ignoring them, though, when we mention them in Q_2 or its negation. The context changes, and our quantifiers take new scopes. Q_1 expresses a truth when uttered before the negation of Q_2, which also expresses a truth, but this is no threat to the duality of the universal and the existential, even though Q_1 and $\sim Q_2$ appear to have the form of $(\forall x)\sim Fx$ and $(\exists x)Fx$.

On my treatment of counterfactuals, safety and sensitivity will be equivalent, when we hold fixed the context set. This is easily seen: a belief that *p* is sensitive just in case $\sim p$ $\square\!\!\rightarrow \sim B(p, E)$. That is, just in case all the not-*p* possibilities are not-B(*p*, E) possibilities; a belief that *p* is safe just in case B(*p*, E) □→ *p*. That is, just in case all the B(*p*, E) possibilities are *p* possibilities. Counterfactuals *do* contrapose, because universals contrapose, and counterfactuals are kinds of universals.

5.5. Sensitivity is consistent with closure.

Sensitivity, understood in terms of my account of counterfactuals, does not imply closure failure for knowledge. Apparent counterexamples equivocate. Here is the alleged consequence of safety I began with:

> I know that I see Laura—if I didn't, it wouldn't appear to me that I did; but I don't know that I don't see a perfect imposter duplicate Laura—if I did, I'd still think that I saw Laura.

The solution is straightforward and familiar: once we are taking seriously the imposter-Laura possibility, it is no longer true to say, 'I know that I see Laura'—and neither is it true to say 'if I didn't see Laura, I wouldn't think that I did.' So there is no counterexample to closure; deductive inference preserves knowledge. 'Knowledge' here is understood univocally; it is true that sometimes we truly say sentences of the form 'S knows that *p*', and also truly say 'S does not know that *q*', even when S has correctly deduced *q* from *p*. These are cases in which our context is shifting between sentences.

5.6. Knowledge entails safety and sensitivity.

I remarked above that, as many authors have observed, sensitivity is an attractive requirement on knowledge. On the proposed accounts of knowledge and counterfactuals, the requirement is easily explained. A belief that *p*, based on E, is sensitive just in case if *p* weren't the case, S wouldn't believe that *p* on the basis of E. That is to say, given my treatment of counterfactuals, a belief is sensitive just in case there are no possibilities in which S believes that *p* because of E, but *p* is false.

Suppose S's belief that *p*, based on E, is insensitive. Then there are some possibilities in which S believes *p* because of E, where *p* is false. Since these are possibilities in which S believes *p* on the basis of E, these are possibilities where E obtains. So it follows directly from insensitivity that there are possibilities in which E and not *p*. On the proposed account of knowledge, S's belief that *p*, based on E, is knowledge only if there are no possibilities in which E and not *p*. So, on the proposed accounts of knowledge and counterfactuals, knowledge entails sensitivity, which is equivalent to safety.

Conclusion

Knowledge and counterfactuals are intimately related; it is no accident that so many attempts have been made to understand the one in terms of the other. My suggestion is to understand both in terms of context-sensitive restricted universal quantifiers; furthermore, the way one belief leads to another when we acquire knowledge by epistemic inference is just the same way one act of imagination leads to another when we come to learn a counterfactual.

A contextualist story about both knowledge and counterfactuals is correct; each inherits its context-sensitivity from the context-sensitive 'all possibilities'. Understanding them this way explains the parallel nature of the skeptical paradox and the parallel challenge against ordinary counterfactuals.

Bibliography

Alexander, J., & Weinberg, J. (2007). Analytic epistemology and experimental philosophy. *Philosophy Compass 2*(1): 56-80.

Augustine. (1999). *The confessions of Saint Augustine.* (E. B. Pusey Trans.) New York: Modern Library.

Bealer, G. (1999). A theory of the a priori. *Philosophical Perspectives, 13*:29-29-55.

Bealer, G. (2002). Modal epistemology and the rationalist renaissance. In T. Gendler, & J. Hawthorne (Eds.), *Conceivability and possibility* Oxford: Clarendon. (Pp. 71-126).

Bisiach, E., & Luzzatti, C. (1978). Unilateral Neglect of Representational Space. *Cortex* 14: 129-133.

Boghossian, P. (2003). Blind reasoning. *Aristotelian Society Supplementary, 77*(1): 225–248.

Brann, E. (1991). *The world of the imagination: Sum and substance*. Rowan & Littlefield.

Budd, M. (1989). *Wittgenstein's philosophy of psychology*. London and New York: Routledge.

Chalmers, D. J. (2002). Does conceivability entail possibility? In T. Gendler & J. Hawthorne (Eds.), *Conceivability and possibility*. Oxford: Clarendon. (Pp. 145-200).

Currie, G. (1990). *The nature of fiction*. Cambridge and New York: Cambridge University Press.

Currie, G. (1991). Work and text. *Mind, 100*: 326-340.

Currie, G. (2000). Imagination, delusion and hallucinations. *Mind and Language, 15*(1): 168-183.

Currie, G. and Ravenscroft, I. (2002). *Recreative minds*. New York: Oxford University Press.

Descartes, R. (1986). *Meditations on First Philosophy*. New York: Cambridge University Press.

Devitt, M. (2006). Intuitions. In V. G. Pin, J. I. Galparaso, and G. Arrizabalaga, eds., *Ontology Studies Cuadernos de Ontologia: Proceedings of VI International Ontology Congress*. (Pp. 169-76).

Edgington, D. (1995). On conditionals. *Mind, 104*(414): 235-329.

Edgington, D. (manuscript). Do counterfactuals have truth conditions? Presented at the 2006 Eastern APA.

Egan, A. (forthcoming). Imagination, delusion, and self-deception. Forthcoming in *Delusions, Self-Deception, and Affective Influences on Belief-formation*, Bayne and Fernandez, eds., Psychology Press.

Elga, A. (2007). Reflection and disagreement. *Noûs, 41*(3): 478-502.

Field, H. (2000). A priority as an evaluative notion. In P. Boghossian, & C. Peacocke (Eds.), *New essays on the A priori*. New York: Oxford University Press. (Pp. 117–49)

Fine, K. (1975). Review of David Lewis's *Counterfactuals*, *Mind, 84*: 451-58.

Foulkes, D. (1999). *Children's dreaming and the development of consciousness.* Cambridge, Mass.: Harvard University Press.

Gendler, T. (1998). Galileo and the indispensability of scientific thought experiment. *British Journal for the Philosophy of Science, 49*(3): 397-424.

Gendler, T. (2000). The puzzle of imaginative resistance. *Journal of Philosophy, 97*(2):

55-81.

Gendler, T. (2002). Personal identity and thought experiments. *The Philosophical Quarterly, 52*(206): 34-54.

Gendler, T. (2003). On the relation between pretense and belief. In M. Kieran, ed., *Imagination, philosophy and the arts*. Routledge. (Pp. 125-141).

Gendler, T. (2006a). Imaginative resistance revisited. In S. Nichols, ed. *The architecture of the imagination*. New York: Oxford University Press. (Pp. 149-174.)

Gendler, T. (2006b). Imaginative contagion, *Metaphilosophy, 3*(2): 183-203.

Gendler, T. (2007). Philosophical thought experiments, intuitions, and cognitive equilibrium. *Midwest Studies in Philosophy, 31*: 68-89.

Gendler, T. (2008). Self-deception as pretense. *Philosophical Perspectives: Mind*, forthcoming.

Gillies, A. (2007). Counterfactual scorekeeping. *Linguistics and Philosophy, 30*(3): 329-60.

Goldman, A. I. (2001). Replies to contributors. *Philosophical Topics, 2*9: 261-511.

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

Goldman, A. I. (2006b). Imagination and simulation in audience responses to fiction. In S. Nichols, ed. *The architecture of the imagination*. New York: Oxford University Press. (Pp. 41-56.)

Goldman, A. I. (2007). Philosophical intuitions: Their target, their source, and their epistemic status. *Grazer Philosophische Studien, 74*: 1-25.

Goodman, N. (1983). *Fact, fiction, and forecast* (4th ed.). Cambridge, Mass.: Harvard University Press.

Hájek, A. (manuscript). Most counterfactuals are false. Unpublished manuscript.

Hanley, R. (2004). As good as it gets: Lewis on truth in fiction. *Australasian Journal of Philosophy, 82*: 112-128.

Hill, C. (2006). Modality, modal epistemology, and the metaphysics of consciousness. In S. Nichols (Ed.), *The architecture of the imagination: New essays on pretence, possibility, and fiction*. New York: Oxford University Press. (Pp. 205-236).

Hjort, M., & Laver, S. (1997) *Emotion and the Arts*. New York: Oxford University Press.

Hobson, J. A. (1999). *Dreaming as delirium: How the brain goes out of its mind.* Cambridge, Mass: The MIT Press.

Horowitz, T. (1998). Philosophical intuitions and psychological theory. *Ethics, 108*: 367-85.

Hume, D. (1978). *A Treatise of Human Nature*, P. H. Nidditch & L. A. Selby-Bigge, eds. New York: Oxford University Press.

Ichikawa, J. (2008). Scepticism and the imagination model of dreaming. *The Philosophical Quarterly* , 58(232): 519-27.

Ichikawa, J. (forthcoming). Dreaming and imagination. *Mind & Language*, forthcoming.

Ichikawa, J. & Jarvis, B. (forthcoming). Thought-experiment intuitions and truth in fiction. *Philosophical Studies*, forthcoming.

Jackson, F. (1998). *From metaphysics to ethics*. Oxford: Oxford University Press.

King, J. (2007). What in the world are the ways things might have been? *Philosophical Studies, 133*(3): 443-453.

Kitcher, P. (1980). A priori knowledge. *The Philosophical Review, 89*(1): 3-23.

Kitcher, P. (2000). A priori knowledge revisited. In P. Boghossian, & P. Benacerraf (Eds.), *New essays on the A priori*. New York: Clarenden Press. (Pp. 65-91).

Kornblith, H. (2003). *Knowledge and its place in nature*. New York: Oxford University

Press.

Kosslyn, S. (1994). *Image and brain*. Cambridge, MA: The MIT Press.

Kosslyn, S., Ganis, G., & Thompson, W. (2006). Mental imagery and the human brain. In Q. Jing., M. R. Rosenzweig, G. d'Ydewalle, H. Zhang, H-C. Chen, & K. Zhang (Eds.), *Progress in Psychological Science Around the World, v. I*. New York: Psychology Press.

Kripke, S. A. (1980). *Naming and necessity*. Cambridge, Mass.: Harvard University Press.

Laberge, S. (2004). *Lucid dreaming: A concise guide to awakening in your dreams and in your life*. New York: Sounds True Press.

Lewis, D. K. (1973). *Counterfactuals*. Oxford: Basil Blackwell Press.

Lewis, D. K. (1978). Truth in fiction. *American Philosophical Quarterly, 15*: 37-46.

Lewis, D. K. (1986). Counterfactuals and time's arrow, in *Philosophical Papers* Volume 2. New York: Oxford University Press. (Pp. 32-66).

Lewis, D. K. (1999). Elusive knowledge. In *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press. (Pp. 418-45).

Malcolm, N. (1959). *Dreaming*. New York: Humanities Press.

Matravers, D. (1997). The paradox of fiction: The report versus the perceptual model. In M. Hjort & S. Laver (Eds.), *Emotion and the Arts*. New York: Oxford University Press.

McCawley, J. E. (1996). Conversational scorekeeping and the interpretation of conditional sentences, in M. Shibatani and S. A. Thompson, eds., *Grammatical constructions: Their form and meaning*. New York: Oxford University Press. (Pp. 77-102).

McGinn, C. (2004). *Mindsight: Image, dream, meaning*. Cambridge, Mass.: Harvard University Press.

Nichols, S. (2004). Imagining and believing: The promise of a single code. *Journal of Aesthetics and Art Criticism, 62*(Special issue on Art, Mind, and Cognitive Science): 129-139.

Nichols, S. (2006). Introduction. In S. Nichols, ed., *The Architecture of the imagination*, New York: Oxford University Press. (Pp. 1-16).

Nichols, S., & Stich, S. (2000). A cognitive theory of pretense. *Cognition, 74*(115-47)

Nichols, S., Stich, S., and Weinberg, J. (2003). Meta-skepticism: meditations in ethno-epistemology. In S. Luper, ed., *The Skeptics*. Aldershot, UK: Ashgate Publishing. (Pp. 227-247.)

Nisbett, R. & Borgita, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 32.

Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology*, 21(4): 422-452.

Priest, G. (1999). Sylvan's box: A short story and ten morals. *Notre Dame Journal of Formal Logic, 38*(4): 573-582.

Pritchard, D. (2005). *Epistemic luck*. New York: Oxford University Press.

Pust, J. (2000). *Intuitions as evidence*. New York and London: Garland Publishing, 2000.

Putnam, H. (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.

Pylyshnyn, Z. (1999). What's in your mind? In E. Lepore & Z. Pylyshnyn (Eds.), *What is Cognitive Science?* Malden, MA: Blackwell Publishing.

Pylyshnyn, Z. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25: 157-182.

Samuels, R. and Stich, S. (2004). Rationality and psychology. In A. Mele, P. Rawling, eds., *The Oxford Handbook of Rationality*. Oxford: Oxford University Press. (Pp. 279-300.)

Sartre, J. (1991). *The Psychology of Imagination*. Secaucus, NJ: Citadel Press.

Schaffer, J. (2001). Knowledge, relevant alternatives and missed clues. *Analysis, 61*(271): 202-8.

Schwitzgebel, E. (2002). Why did we think we dreamed in black and white? *Studies in History and Philosophy of Science*, 33(4): 649-660.

Skolnick, D., & Bloom, P. (2006a). The intuitive cosmology of fictional worlds. In S. Nichols (Ed.), *The architecture of the imagination: New essays on pretense, possibility, and fiction*. New York: Oxford University Press. (Pp. 73-73-86).

Skolnick, D., & Bloom, P. (2006b). What does batman think about SpongeBob? Children's understanding of the fantasy/fantasy distinction. *Cognition, 101*(1): B9-B18.

Solms, M. (1997). *The neuropsychology of dreams*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Solms, M. & Turnbull, O. (2002). *The brain and the inner world*. New York: Other Press, LLC.

Sorensen, R. (1996). Modal bloopers: Why believable impossibilities are necessary. *American Philosophical Quarterly*, 33(1): 247-61.

Sosa, D. (2006). Scenes seen. *Philosophical Books*, 47(4): 314-325.

Sosa, E. (1999). How to defeat opposition to Moore, *Philosophical Perspectives*.

Sosa, E. (2002). Reliability and the a priori. In T. Gendler, & J. Hawthorne (Eds.), *Conceivability and possibility*. New York: Oxford University Press. (Pp. 369-384).

Sosa, E. (2007a). *A virtue epistemology: Apt belief and reflective knowledge, volume I.* New York: Oxford University Press.

Sosa, E. (2007b). Experimental philosophy and philosophical intuition. *Philosophical Studies, 132*(1): 99-107.

Sosa, E. (forthcoming). A defense of the use of intuitions in philosophy, in D. Murphy, ed., *Stich and his critics*, Malden, MA: Blackwell.

Spivey, M., Tyler, M. Richardson, D., & Young, E. (2000). Eye movements during comprehension of spoken scene descriptions. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Mahwah, NL: Erlbaum. (Pp. 487-492)

Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.

Stanley, J. (2005). *Knowledge and practical interests*. New York: Oxford University Press.

Stanley, J. (forthcoming). Knowledge and certainty. *Philosophical Issues*, forthcoming.

Stanley, J. & Williamson, T. (2001). Knowing how. *The Journal of Philosophy, 98*(8): 411-44.

Stich, S. (1990). *The fragmentation of reason*. Cambridge, MA: The MIT Press.

Stich, S. (forthcoming). Reply to Sosa, in D. Murphy, ed., *Stich and his critics*, Malden, MA: Blackwell.

Stock, K. (2003). The tower of Goldbach and other impossible tales. In M. Kieran, & D. Lopes (Eds.), *Imagination, philosophy and the arts.* Routledge.

Swain, S., Alexander, J., & Weinberg, J. (2008). The instability of philosophical intuitions, *Philosophy and Phenomenological Research, 76*(1): 138-55.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review 90*(4).

Tye, M. (1995). *The problems of consciousness*. Cambridge, MA: The MIT Press.

Walton, K. L. (1990). *Mimesis as make-believe: On the foundations of the representational arts*. Cambridge, Mass.: Harvard University Press.

Walton, K. L. (1997). Spelunking, simulation, and slime: On being moved by fiction. In M. Hjort, & S. Laver (Eds.), *Emotion and the Arts*. New York: Oxford University Press. (Pp. 37-49).

Wason, P., and Johnson-Laird, P. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Philosophy*, 61: 509-15.

Weatherson, B. (2004). Morality, fiction, and possibility. *Philosopher's Imprint 4*(3): 1-27.d

Weinberg, J. (2007). How to challenge intuitions empirically without risking skepticism. *Midwest Studies in Philosophy, 31*(1): 318-43.

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

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

Williamson, T. (2004). Philosophical 'intuitions' and scepticism about judgement. *Dialectica, 58*(1), 109-153.

Williamson, T. (2005). Armchair philosophy, metaphysical modality and counterfactual thinking (presidential address). *Proceedings of the Aristotelian Society, 105*(1): 1-23.

Williamson, T. (2007). *The Philosophy of Philosophy*. Malden, MA: Blackwell Publishing.

Wittgenstein, L. (1970). *Zettel*. G. E. Anscombe, G. H. Von Wright (Eds.), G. E. Anscombe (Trans.). Berkeley, CA: University of California Press.

Yablo, S. (1993). Is conceivability a guide to possibility? *Philosophy and Phenomenological Research, 53*(1): 1-42.

Yablo, S. (2002). Coulda, woulda, shoulda. In T. Gendler, & J. Hawthorne (Eds.), *Conceivability and possibility*. Oxford: Clarendon. (Pp. 441-492).

Curriculum Vita: Jonathan Ichikawa

Education:

Dates	*School*	*Degree*
1999-2003	Rice University	B.A., Philosophy and Cognitive Science (2003)
2003-2005	Brown University	M.A., Philosophy (2005)
2005-2008	Rutgers University	Ph.D., Philosophy (2008)

Publications:

2008:

"Scepticism and the imagination model of dreaming," *The Philosophical Quarterly* , 58(232): 519-27.

Forthcoming:

"Thought-Experiment Intuitions and Truth in Fiction," with Benjamin Jarvis, forthcoming in *Philosophical Studies*.

"Dreaming," with Ernest Sosa, forthcoming entry in the *Oxford Companion to Consciousness.*

"Dreaming and Imagination," forthcoming in *Mind & Language*.

Teaching Experience:

Philosophy 11: The Nature of Fiction, Brown University (Teaching Fellowship), Fall 2005

Philosophy 101: Reasoning and Value, Bridgewater State College (Visiting Lecturer), Spring 2006

Philosophy 913: Ethics: Right and Wrong, Summer @ Brown (Three-week intensive summer course for high school students), Summer 2006