UNIVERSITY OF CALIFORNIA
RIVERSIDE

Armchair Knowledge and Modal Skepticism: A Rapprochement

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Philosophy

by

Felipe Edwin Leon

June 2009

Dissertation Committee:
      Dr. Peter J. Graham, Chairperson
      Dr. Peter Kung
      Dr. Michael Nelson
      Dr. Eric Schwitzgebel

UMI Number: 3374404

Copyright 2009 by
Leon, Felipe Edwin

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy
submitted.   Broken or indistinct print, colored or poor quality illustrations
and photographs, print bleed-through, substandard margins, and improper
alignment can adversely affect reproduction.
In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted.   Also, if unauthorized
copyright material had to be removed, a note will indicate the deletion.

UMI®

Copyright by
Felipe Edwin Leon
2009

Signature Page on file in the

Graduate Division

ACKNOWLEDGEMENTS

The members of my dissertation committee are a wonderful group of people to work with. Thanks to Peter Kung at Pomona College. Peter was extremely generous with his time, and his incisive comments, friendliness, and encouragement have made a tremendous impact on the quality and substance of my dissertation. Thanks also to Eric Schwitzgebel. From my first seminar to my Master's thesis, and on through the dissertation process, Eric's instruction, guidance, encouragement, and support have been indispensable. I'd also like to thank Michael Nelson for his gracious spirit and his penetrating comments on the core ideas of my dissertation. I'm grateful for the opportunity to have worked with him. Thanks especially to my adviser, Peter J. Graham. Peter has been my mentor throughout my time in the Philosophy graduate program at UC Riverside. As an adviser and as a person, he is without match.

UC Riverside's Philosophy department is a congenial and stimulating environment for philosophical development. Thanks to Paul Hoffman for opening my eyes to the subtlety and depth of Descartes' work. It was Hoffman's seminar on Descartes that sparked what would become a deep and abiding interest in modal epistemology in particular and philosophical methodology in general. It was also nice to have a fellow surfer in the department! Thanks also to Erick Reck and Charles Siewert for their comments and advice during the early stages of the dissertation process. And thanks to John Martin Fischer and Howard Wettstein for their instruction,

encouragement, and guidance at various stages in my graduate school career. Finally, I'd like to thank Kathy Saylor for all her help and support throughout my time in the program.

I'm also grateful to my fellow graduate students for their friendship, camaraderie, and philosophical discussion. My thanks to Josh Bright, Jacob Affolter, Jessica Ludescher, Jeff Yim, Rachel Jenkins, Dan Speak, Rico Vitz, Matt Talbert, Rich Doan, Koki Funato, Nick Baiamonte, and R. Dillon Emerick. Thanks especially to Neal Tognazzini. Neal is a good friend, a great interlocutor, and a first-rate person. I'm especially grateful for his encouragement, discussion, and advice while developing some core ideas in the second half of my dissertation. Our discussions led to our collaboration on a paper we presented at the 2007 APA Pacific Division meeting. We later revised it and submitted it for publication.

Relatedly, I thank our audience at said conference for their comments (especially Kevin Timpe and Michael McKenna). Thanks also to *Philosophy and Phenomenological Research* for their kind permission to include amended and non-amended portions of our forthcoming article, "Why Frankfurt-Examples Don't Need to Succeed to Succeed" (with Neal Tognazzini) in chapters 5 and 6 of my dissertation.

Help in writing my dissertation traces back to my years as an undergraduate. Thanks to Jill Buroker, Susan Finsen, Chris Naticchia, Tom Moody, Beverly Gallo of the Philosophy faculty at CSU San Bernardino for their excellent mentoring. Their teaching and encouragement put me in a position to pursue philosophy at the graduate level. I'd

especially like to thank Tony Roy for his excellent and tireless mentorship, both inside the class and out.

Many thanks are due to my friends, Chris and Sonia Ward, Tanya and Kevin Turner, Heidi and Dave Matthewman, Scott Collins, Bob Edwards, Katie Battino, Dave Silva, Jason Hackerd, my sister-in-law, Vivian Mooers, and my brothers-in-law, Erick Romero and Russ Haley for their love and support. Thanks especially to my best friend, Andy Battino. Few are lucky enough to have a friend as close as a brother. His friendship and encouragement have sustained me through good times and bad.

Finally, I'd like to thank my family. My mother and father, Mel and Donna Dresp, provided me with the love and encouragement that gave me the confidence to pursue graduate studies in Philosophy. Thanks also to my sister, Elena, my grandparents, Ed and Julie Hanson, and my uncle, Frank Hanson, for their encouragement. I'd also like to thank my mother-in-law, Teresa Romero, and my father-in-law, Fabio Romero, for their love and encouragement. Thanks especially to my wonderful wife, Soraya. Her unyielding love, support, and encouragement sustained me throughout my years of grad school. Finally, I'd like to thank my daughter, Abby, for the love and support she expresses in her own special ways. Daddy is turning off the computer to come and play.

DEDICATION

For Soraya and Abby

ABSTRACT OF THE DISSERTATION

Armchair Knowledge and Modal Skepticism: A Rapprochement

by

Felipe Edwin Leon

Doctor of Philosophy, Graduate Program in Philosophy
University of California, Riverside, June 2009
Dr. Peter J. Graham, Chairperson

Thought experiments play an indispensable role in philosophical methodology. One

wonders, for example, how philosophers could come to think that knowledge isn't

justified true belief, that "meanings ain't in the head", and that moral responsibility

doesn't require alternative possibilities, without the use of thought experiments. Despite

this, however, the legitimacy of thought experiments is at least initially puzzling: one

would think that significant knowledge of the world requires extensive empirical

investigation. But since thought experiments are done from the armchair, how can they

tell us about the world?

A standard account of the nature and utility of thought experiments provides an

answer to this question, and in a way that fits naturally with a standard picture of the facts

philosophers investigate: Philosophers investigate the essences of things. But a thing's

essential properties are *modal* properties. Thus, one can discern a thing's essence by discovering its modal profile. But thought experiments shed light on modal facts. Therefore, thought experiments are naturally suited as tools for the armchair philosopher.

In my dissertation, I argue that the standard account is inadequate, and sketch a more promising account. First, I argue that our knowledge of possibility is restricted to the relatively humdrum, and if so, then since the standard account ties the utility of thought experiments to our knowledge of possibility, too many thought experiments will be ruled out as useless.

Second, I explicate and defend a hitherto-underappreciated type of thought experiment, which I call the *non-modal thought experiment*. These shed light on facts about entities that are finer-grained than those capturable by their modal profiles. I argue that non-modal thought experiments often succeed at just the points where the more familiar modal thought experiments fail, and thus that the two are naturally suited to complement one another in the philosopher's practice.

Finally, I sketch a more nuanced account of the variety and utility of thought experiments. I then illustrate its virtues, such as its ability to account for the epistemic force of a wider range of thought experiments, and its utility in getting beyond dialectical impasses caused by conflicting modal intuitions.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

1. Locating the Topic

Consider the following variation on a familiar story[1]. A philosopher considers the

traditional analysis of knowledge:

> (TAK) S knows that p $=_{df}$ (i) S believes that p, (ii) p is true, and (iii) S is
>
> justified in believing that p.

Trying to determine whether TAK is true, he does his best to come up with a

counterexample. Unaware of any *actual* case that would do the trick, he tries to imagine

a scenario according to which someone has a justified true belief that isn't knowledge.

Somewhat shocked, he succeeds: he imagines a scenario in which Smith comes to

believe, on the basis of excellent evidence, the false belief that Jones owns a Ford. From

this, Smith validly deduces the proposition that Jones owns a Ford or Brown is in

Barcelona. Since this belief logically follows from the former belief, he comes to

justifiedly believe the disjunctive statement as well. As it turns out -- and unbeknownst

to Smith -- Brown *is* in Barcelona. Thus, Smith has a justified true belief that Jones owns

a Ford or Brown is in Barcelona. But since it's just a lucky accident that Smith

justifiedly believes something true, it's not a case of knowledge. On the basis of this

---

[1] Gettier, Edmund. "Is Justified True Belief Knowledge?", *Analysis* 23 (1963), pp. 121-123.

1

imagined scenario, the philosopher comes to reject TAK. In fact, it seems that he's *justified* in doing so.

The preceding case introduces the three related topics of my dissertation. First, the case involves a *thought experiment*. That is, in the case, a philosopher considers a scenario that he takes to represent an actual or possible state of the non-imaginary world. Second, the philosopher seems to assume that his counterexample scenario is metaphysically possible. That is, he seems to assume the truth of some substantive, positive account of *modal epistemology*. Finally, insofar as the philosopher uses the thought experiment with the intent of gaining knowledge or justified beliefs about the nature of the world, he's engaged in *armchair philosophy*. My dissertation focuses on these three things -- thought experiments, modal epistemology, and armchair philosophy. In particular, I develop an account of the variety and utility of thought experiments. This, in turn, will involve an account of the extent to which thought experiments involve and are committed to a substantive modal epistemology. And finally, my account of these matters will hopefully illuminate, and support the viability of, armchair philosophy. But before I can meaningfully give more specific statements of my main theses, I will need to do some further stage-setting.

2. Armchair Philosophy: A Rough Characterization

Although the following is admittedly a very general and highly idealized characterization of the strand of philosophy I aim to pick out, it will nonetheless prove helpful for our purposes to individuate what I am calling 'armchair philosophy' by means of the

2

following three theses. The first is that philosophy at least partly relies on the utility of thought experiments, in that the thought experiment is an indispensable component of philosophical methodology. Just as the scientist uses empirical experiments to discern the nature of the objects of her study, so the philosopher uses thought experiments to discern the nature of the objects of *her* study. And just as the scientist would no longer be able to do science if she couldn't utilize empirical experiments, so (according to the first thesis) the philosopher would no longer be able to do philosophy if she could not utilize thought experiments. One wonders, for example, how philosophers could come to accept that justified true belief isn't knowledge, that reference isn't determined by an expression's associated description, that moral responsibility doesn't require the ability to do otherwise, and many other philosophical theses, without the use of thought experiments.

The second thesis is that thought experiments have utility only if they reveal facts about the nature of things in the world. In particular, their utility depends upon their ability to help us distinguish between an entity's essential properties from its accidental properties. If thought experiments aimed to tell us about the nature of the world and its constituents and yet failed miserably, or if they were only able to tell us about our *concepts* about the world, then (according to the second thesis) they would have no legitimate role in philosophy.

Finally, the third thesis is that thought experiments can reveal facts about the world and its constituents only by means of revealing *modal* facts about them. According to this thesis, legitimate thought experiments aim to accurately represent the way the world and its constituents can't be, could be, or must be. This aim, assuming they can

3

achieve it, makes thought experiments naturally suited as tools of the armchair philosopher for discerning the nature of things. For if essential and accidental properties are *modal* properties, then discovering a thing's essence amounts to discovering its modal profile. Thus, if they're effective in achieving this aim, then they're suitable as tools for the armchair philosopher. But conversely, if they fail to do this, they are useless with respect to telling us about the essence of things in the world.

To sum up: a large and historically important strand of philosophy can be individuated in virtue of three theses: (i) that philosophy at least partly depends on the utility of thought experiments; (ii) that the utility of thought experiments depends on their ability to reveal the nature of the world; and (iii) that their ability to reveal the nature of the world depends on their ability to reveal modal facts about the world. And we are calling approaches to philosophy that embody at least these three theses, *Armchair Philosophy*.

3. The Challenge: Modal Skepticism and BAAAP

It goes without saying that a good many philosophers reject Armchair Philosophy. While a variety of reasons have been given, a significant number of them share a common theme, viz., that we lack an ability to reliably discover modal facts about the world – at least from the armchair. So, for example, we've said that according to thesis (iii), the legitimacy of thought experiments depends on their ability to reveal facts about things at other possible worlds. However, it's hard to see how sitting in one's armchair and imagining various scenarios could be a reliable way of getting at such facts if the latter

4

are really mind-independent. For knowledge of mind-independent reality would seem to require causal contact. However, unlike ordinary perception, it's hard to see how thought experiments – or anything else, for that matter -- could put us in causal contact with *non-actual* facts about the world.

But even waiving this problem, there is the worry that the source of our modal seemings has been shown to be unreliable, or at least called into serious doubt. For many necessary truths about the world are only discoverable *a posteriori* -- away from the armchair, so to speak.[2] So, for example, Hesperus is the same heavenly body as Phosphorus, and so Hesperus is *necessarily* Phosphorus (assuming the necessity of identity), and yet it required empirical investigation to discover this identity. Furthermore, there is reason to think that what holds here for individuals holds also for natural kinds. Thus, water is the same thing as H20, in which case water is necessarily H20 (assuming the necessity of identity), and yet it required empirical research to make this discovery. But if these things are so, then it seems that I can imagine Hesperus existing without Phosphorus, when in fact the state of affairs imagined is impossible, as Hesperus *is* Phosphorus. Similarly, a person from an earlier era, or a suitably ignorant person from the present era, could arguably imagine water existing at a world in which H20 does not. But since water *is* H20, this is impossible. Thus, for these sorts of reasons,

---

[2] The canonical texts on this point are Kripke, Saul. *Naming and Necessity* (Cambridge, Mass: Harvard University Press, 1980); and Putnam, Hilary. "The Meaning of 'Meaning'", in *Mind, Language and Reality: Philosophical Papers, Vol. 2* (Cambridge: Cambridge University Press, 1975), pp. 215-271.

and many more[3], there are non-trivial concerns about the reliability of the source(s) of our modal beliefs.

The concerns about the quality of our sources for modal information generate a significant challenge for Armchair Philosophy. We can express the connection between the challenge to modal knowledge and the challenge to Armchair Philosophy by means of the following argument, which I will call *The Basic Argument Against Armchair Philosophy* (BAAAP):

> **The Basic Argument Against Armchair Philosophy (BAAAP)**
> 1. Armchair philosophy is legitimate only if thought experiments give us insight into the nature of things.
> 2. Thought experiments give us insight into the nature of things only if they justify modal claims.
> 3. Thought experiments don't justify modal claims.
> 4. So, thought experiments don't give us insight into the nature of things.
> 5. So, armchair philosophy is not legitimate.

4. Three Standard Responses: Two Concessive; One Not

What to make of BAAAP? Well, it's valid. Furthermore, (4) follows from (2) and (3), and (5) follows from (1) and (4). So, the argument turns on (1), (2), and (3). Thus, leaving aside cases of agnosticism and cases of abandonment of the philosophical enterprise, there are four basic ways to react to BAAAP: one can accept all the premises, and thus reject armchair philosophy, or one can reject (1), (2), or (3). Now suppose we idealize away from actual practitioners, and associate possible responses to BAAAP with general metaphilosophical stances. The following characterization of the camps corresponding to these reactions is *very* rough, and doesn't represent cases of overlap, but

---

[3] We'll discuss a representative sampling of the problems for modal knowledge in chapters 1-3, as well as appendices I and II.

our purposes only require that we carve up the metaphilosophical landscape in terms of very general and coarse-grained distinctions.

It's interesting to note that, to date, only three of the four possible responses to BAAAP seem to have been taken up. Of these, the first two are concessive, and the third is non-concessive. Let's start with the former two. Thus, one possible reaction is to accept the argument as sound and thus reject Armchair Philosophy for some empirically-based approach. There are a variety of ways in which such a response could be spelled out, but for now, we can characterize them loosely and lump them into one camp, which we will call *The Naturalized Philosophy Response*.

A second possible reaction is to reject (1). Variations of this response accept the use of thought experiments to evaluate philosophical claims, but concede that we must give up Armchair Philosophy's aim of telling us about the mind-independent world. There are many ways that this approach has been spelled out, but they have in common the view that philosophy is about the task of elucidating our concepts and conceptions of the world and its constituents. We shall therefore call the (idealized) camp corresponding to this response *The Conceptual Analysis Response.*

The third possible reaction – the non-concessive response among the three -- is to reject (3). Versions of this approach adhere to all three theses of Armchair Philosophy. Thus, they're similar to those in The Conceptual Analysis Response in virtue of their conception of philosophy as an armchair discipline. But they differ from them in virtue of their belief that thought experiments give us insight into the nature of reality, and by means of revealing modal facts about it. Call this approach *The Modal Epistemology*

7

Response.

In short, BAAAP poses a challenge to the legitimacy of Armchair Philosophy, and the ways in which philosophers have (in effect) responded to it correspond to the basic sorts of approaches to philosophical practice that have arisen to date: The Naturalized Philosophy Response, The Conceptual Analysis Response, and the Modal Epistemology Response. The Naturalized Philosophy Response accepts BAAAP as sound; the Conceptual Analysis Response rejects (1); and the Modal Epistemology Response rejects (3). The first two camps agree that the gap between our folk concepts and the world can't be bridged, but they disagree about what philosophers should do in light of this fact: One opts for holding on to Armchair Philosophy's methodology while abandoning the quest for knowledge of the world by restricting its scope to the analysis of concepts; the other opts for holding on to the quest for knowledge of (or adequate theorizing about) the world, but abandons armchair methodology for empirical methods, or the results thereof. And The Modal Epistemology Response argues that the modal link remains intact between our concepts and the world, and thus the traditional conception of armchair philosophy -- whereby conceptual analysis and thought experiments yield justified beliefs about the world -- retains its legitimacy.

5. Crisis! The Failure of the Standard Non-Concessive Response

It is at this point in the dialectic that the two main theses of my dissertation make their appearance. *First*, I will argue that, to date, The Modal Epistemology Response is a failure, and further, that the prospects for discovering a plausible account of substantive

8

knowledge of metaphysical possibility are bleak. *Second*, I will argue that despite this, a plausible, minimally-concessive response to BAAAP is available to the armchair philosopher. The latter thesis can be broken down into two. The first asserts that despite our lack of knowledge with respect to the exotic modal claims involved in a number of philosophical thought experiments, we nonetheless have a modest amount of knowledge regarding possibilities relevantly similar to what we experience in the actual world. And, perhaps surprisingly, a considerable amount of philosophical use can be gotten from thought experiments grounded in such knowledge.

The second part relies on the hitherto-unexplored fourth option in responding to BAAAP noted above. To elaborate a bit, recall that BAAAP follows from three basic premises. Now we've seen that the standard responses to date either accept all the premises, reject (1), or reject (3). Thus, interestingly, all the camps at least *de facto* accept (2). But as I will argue, (2) isn't obviously true. And if not, then there's room for a different, minimally concessive response to BAAAP – i.e., a form of armchair philosophy that at least partly relies on thought experiments that have the following two features: (i) they tell us about the world, but (ii) they do so without reference to *modal* facts. To give it a name, we will call this approach, *The Non-Modal Thought Experiments Response*. These, then, are the primary aims of my dissertation.

Before moving on to a summary of the chapters of my dissertation, I should say a brief word about neighboring claims I will not defend here. In particular, I will assume the truth of (1), and thus won't argue against the responses to BAAAP embodied in The Naturalized Philosophy Response and The Conceptual Analysis Response. One must

9

pick one's battles, and I leave these battles for another day in order to achieve my main objective, which is to determine whether, and to what extent, traditional armchair philosophy can survive the challenge of modal skepticism. I will therefore leave discussion of these two broad metaphilosophical stances to the side, and proceed under the following assumption, which I will call 'The Master Claim':

The Master Claim: Thought experiments tell us about the world.

Assuming the truth of The Master Claim, then, the three main responses to BAAAP reduce to two: the Modal Epistemology Response and the Non-Modal Thought Experiments Response. One can then think of my project as exploring the viability of these two responses. Despite the assumption of the Master Claim, I believe that those who fall within the camps corresponding to the Naturalized Philosophy Response and the Conceptual Analysis Response will nonetheless find interest in my project, if for no other reason than to see what can be said on behalf of Armchair Philosophy in the face of the challenge of modal skepticism.

6. Summary of the Chapters

The first three chapters develop and assess the case for premise (3) of BAAAP. In Chapter 1, I argue for a mitigated form of modal skepticism: we have knowledge of humdrum metaphysical possibilities (e.g. that my daughter's ball can get stuck on my roof, that my car can be painted pink, etc.), but we lack knowledge of metaphysical possibilities remote from ordinary experience (e.g., that Putnam's twin water is metaphysically possible). First, I briefly argue for a minimal condition of adequacy for

10

any plausible account of our knowledge of possibility: it must entail that at least humdrum modal claims (e.g., that my daughter's ball can get stuck on my roof, that my car can be painted pink, etc.) are justified. Second, I offer evidence that our judgments about exotic possibilities are unreliable: when we check our judgments about exotic possibilities with respect to things that exist in our world against the backdrop of our scientific understanding of it, it turns out that a significant range of such judgments don't fare so well. Third, I present a dilemma for standard accounts of modal epistemology: either they're so permissive that they admit too many false positives, or they're so restrictive that they entail that even modestly non-humdrum modal claims (e.g., that twin water is metaphysically possible) are unjustified. Fourth, I sketch Peter Van Inwagen's version of mitigated modal skepticism as presented in his paper, "Modal Epistemology", and offer a criticism that undercuts it. Briefly, the problem is that while he draws the boundaries between justified and unjustified modal claims in plausible places, he offers no substantive account of the *sources* of justification for the modal claims he accepts. This leaves his mitigated form of modal skepticism open to the charge of being unprincipled and arbitrarily selective. Finally, I sketch a positive account of the sources of justification for our beliefs about what's metaphysically possible. The most salient source here is *our folk theory of how the world works*, which includes a folk physics, a folk psychology, etc. I argue that the account has many virtues that make it attractive: it explains, in a natural way, our knowledge of humdrum possibilities; it explains the epistemic force of paradigm cases of successful modal thought experiments (e.g., Gettier's thought experiments); it explains common disagreement with respect to claims

11

about exotic possibilities; and it avoids the criticism that undercuts Van Inwagen's account.

In Chapters 2 and 3, I complete my *prima facie* case for mitigated modal skepticism by examining and critiquing prominent contemporary accounts of modal epistemology. Chapter 2 exposits and critiques Stephen Yablo's account, and Chapter 3 exposits and critiques David Chalmers' account. I argue that both accounts suffer from two basic problems: (i) they entail that possibility-claims remote from ordinary experience are unjustified, and thus that their accounts are equivalent to my account in their consequences; and (ii) since they make no advance over my account, their core thesis of imagination as an independent source of modal justification is undercut.

Chapter 4 is a transitional chapter. Here, I briefly draw some important metaphilosophical morals from the first three chapters. It is here that the tension for Armchair Philosophy caused by the challenge of modal skepticism reaches its highest point. For here I use the case for modal skepticism in chapters 1-3 to support the crucial third premise of BAAAP, with the result that an unqualified endorsement of The Modal Epistemology Response appears to be off the table. The result is that Armchair Philosophy has lost its main traditional response to BAAAP. I then suggest a point I develop in the next two chapters: that an essentially non-concessive response to BAAAP remains for the armchair philosopher to adopt: an account of armchair philosophy that partly relies on thought experiments that have the following two features: (i) they can give us information about the nature of reality – i.e., they can reveal *metaphysical* facts –

and (ii) they can achieve this aim without giving us information about possible worlds – i.e., without revealing *modal* facts.

In chapter 5, I motivate, sketch, and defend just such a category of thought experiment: what I call *non-modal* thought experiments. I then raise and address two main obstacles to non-modal thought experiments: possible worlds accounts of strict conditionals, and possible worlds accounts of essential properties. These are obstacles because it seems that philosophers make claims about the nature of things, and that such claims are properly cashed out in terms of necessary truths, viz., in terms of possible worlds analyses of strict conditionals and essential properties. But if so, then it would seem that the only effective way to evaluate and defeat such claims is via appeal to possible worlds at which such claims are false. But since non-modal thought experiments do not make crucial reference to possible worlds, it seems that they're irrelevant to evaluating philosophical claims. I offer two main responses. First, I briefly argue that possible worlds analyses of strict conditionals and essential properties are inadequate. Second, I sketch alternative – or at least surrogate – accounts of strict conditionals and of essence that make no crucial reference to possible worlds. The main significance of this chapter is that it provides the basis for a qualified rejection of premise (2) of BAAAP. For there are a significant number of philosophical claims that can be evaluated via non-modal thought experiments. In these ways, the chapter makes a preliminary exploration of a new sort of response to BAAAP mentioned above, viz., The Non-Modal Thought Experiments Response.

13

In chapter 6, I synthesize the relevant material from the previous chapters to offer an account of the nature and utility of thought experiments. This consists in providing a typology of the various sorts of thought experiments, and the conditions under which each sort is most relevant and useful. Once the framework is laid out, I attempt to illustrate it and demonstrate its utility by applying it to example cases. In particular, I show how to use the framework to clarify and help get beyond philosophical debates where there are deeply entrenched impasses. Since the most novel and unfamiliar aspect of my framework is the non-modal thought experiment, I finish the chapter by showing how such thought experiments can be used to evaluate two further example theses: The Principle of Alternate Possibilities (PAP), and modal analyses of essential properties.

I conclude the chapter by returning to the challenge raised at the beginning of the dissertation, viz., the challenge of BAAAP brought on by our case for modal skepticism. I then attempt to address the challenge in the light of the resultant framework. Briefly, I argue that while our knowledge of possibility is limited to the humdrum, this has little effect on the legitimacy of Armchair Philosophy. For modal skepticism, when properly spelled out and shortcomings are redressed, allows for the legitimacy of modal thought experiments that involve humdrum possibilities, and significant philosophical work can be accomplished with them. And while more exotic modal thought experiments are an illegitimate means of access to facts about the nature of things, there are also exotic *non-modal* thought experiments, and these are a legitimate means of access to such facts. And if these things are so, then modal skepticism does not undermine the legitimacy of

14

Armchair Philosophy, and so we have a response to BAAAP that leaves Armchair Philosophy largely intact.

There are also three appendices. In Appendix I, I critically explore Descartes's modal epistemology. After developing and defending his account against a number of challenges, I argue, first, that his account crucially relies on a number of his own metaphysical assumptions, and second, that while it's a reasonable and useful account of modal epistemology on those assumptions, as a matter of fact they have been shown to be false.

Peter van Inwagen's writings contain the materials for a powerful critique of many of the popular accounts of modal epistemology. Unfortunately, there is no single piece of his where he draws all of his various strands of argumentation together and develops them. Even in his most recent paper on the topic -- "Modal Epistemology"[4] -- his case is more akin to a series of "sketches of landscapes" than a systematic critique of the various accounts of modal knowledge and justification on offer. This lack of clarity and systematicity is evidenced in the lack of agreement as to what his main arguments even are among recent articles addressing his modal skepticism. I try to redress this in Appendix II by developing and systematizing Van Inwagen's defense of modal skepticism.

In Appendix III, I sketch the two-dimensionalist framework that underlies Chalmers' modal epistemology discussed in Chapter 3.

---

[4] *Philosophical Studies* 92 (1998), pp. 67-84.

15

Chapter 1:
Our (Limited) Knowledge of Possibility

> "An important part of education is the acquiring of a sense of the possible; it comes slowly, with difficulty, and it must be nourished by a feeling for inter-relation, the knowledge that in nature one can never do only one thing. Talk of 'logical possibility' may weaken this sense."

> "Locke said that there is no connection between the colour of gold and its malleability, but Locke was wrong."

> -George Seddon[5]

1.1 Introduction

We're all familiar with Putnam's Twin Earth thought experiment[6]:

> **Twin Earth**:
> On Twin Earth, everything is just as it is on Earth. Everything, that is, except for one thing: the stuff that they call *water* – the clear, drinkable stuff that flows in rivers and streams, falls from clouds, freezes at 32 degrees Fahrenheit, etc. – is not composed of H20, but rather some alien substance, XYZ. Thus, when uttered by those on Earth, 'water' refers to H20, and when uttered by those on Twin Earth, 'water' refers to XYZ.

This thought experiment has convinced many that the meaning of natural kind terms is not, or at least not exhausted by, its descriptive content. However, at least *prima facie*, the Twin Earth thought experiment is a success only if twin water – i.e., an alien substance that perfectly plays the watery role – is metaphysically possible. Now no one but the most radical sort of modal skeptic would deny that we have at least some

---

[5] "Logical Possibility" *Mind* 81(Oct., 1972), pp. 481-482, and p. 486, respectively.

[6] The following is a paraphrase of his version as found in "The Meaning of 'Meaning'", in *Mind, Language and Reality: Philosophical Papers, Vol. 2* (Cambridge: Cambridge University Press, 1975), pp. 215-271. An initial statement of the thought experiment occurs on pp. 223-224.

knowledge of metaphysical possibility. Thus, I know that my table can be moved about, that my car can be painted a different color, and that my daughter's ball can get stuck on my roof. Both you and I know these things, and many similar things as well, to be metaphysically possible. However, I will argue that despite what Putnam has said, and despite the many advances in modal epistemology, *nobody knows whether twin water is metaphysically possible.*

In this chapter, I will begin the task of arguing for this claim. Toward that end, I'll offer some initial reasons for thinking that our knowledge of metaphysical possibility is limited, in that we don't know whether such exotic things (or thing-candidates) as twin water are metaphysically possible. I will then consider an account of the nature and extent of our modal knowledge that accords with these considerations: Peter van Inwagen's mitigated version of modal skepticism. Next, I will offer reasons for thinking that his account is inadequate as stated. Following that, I will sketch a positive account of our knowledge of possibility that avoids the weaknesses in van Inwagen's account, but that is (perhaps unfortunately) even more far-reaching in its skeptical implications. Our discussion here will provide the fundamental lines of reasoning employed in chapters 2 and 3, where I will argue that, perhaps surprisingly, the major extant sophisticated accounts of our knowledge of possibility are roughly equivalent in their implications to my account. I conclude the chapter by bringing our preliminary investigation to bear on our larger project of exploring the connection between thought experiments and the world.

17

1.2 "Why So Skeptical?" Seddon and the Little Leeway Thesis

I said above that while we may well know many relatively humdrum possibility claims,

our knowledge of metaphysical possibility doesn't extend much farther beyond them.

Bold words! What do I have to say for myself? While I have a lot of explaining to do, I'd

like to start off with a consideration that I think is particularly weighty. For it seems that

we can check the reliability of at least some of our possibility-judgments against the

backdrop of our scientific understanding of the world. And as it turns out, a significant

range of such judgments doesn't fare so well. So, for example, George Seddon[7] has

argued that some basic features of scientific theories pose a serious challenge to a naïve

view of our knowledge of possibility. For as he points out, our well-developed and well-

confirmed scientific theories not only reveal the nature of objects and events in the world,

but they also reveal some facts about the fundamental structure of reality that are at odds

with the sorts of exotic possibilities philosophers are often interested in. For, first, they

tell us that the features of things in the world are *essentially interdependent and*

interrelated, in that kinds are often a part of an interlocking, nested hierarchy of entities

(in some cases, e.g., in terms of species, genus, class, order, etc.), and that the alteration

of one feature of a thing requires an alteration in others, which in turn requires a change

in others, all the way down to the categorical properties at the base: a change in even the

"node" features often requires a change in the "root" features, so to speak.

Another fact that our scientific theories reveal about the structure of reality is that

natural kinds occupy *highly-specific niches*, in that a given natural kind is typically

---

[7] "Logical Possibility", *Mind* 81(Oct.,1972), pp. 481-492.

18

"sandwiched" between two neighboring kinds, such that fairly minor alterations to a member of a kind would either make it go out of existence *qua* that thing, or land it in another natural kind. Given these well-supported facts about things and kinds – interrelatedness of kinds, interdependence of features of a given kind, and highly-specific niches for each kind – it follows that there is little leeway for things to be altered without either making them cease to exist *qua* that thing or kind of thing, or otherwise requiring widespread, perhaps global, changes in the world's structure and its history. Call this *The Little Leeway Thesis*:

> *The Little Leeway Thesis*: The structure of the world leaves little leeway
>
> for altering its occupants.

The Little Leeway Thesis poses a problem for exotic[8] thought experiments. For such thought experiments represent things and states of affairs significantly unlike those we find in the actual world, or at least radical alterations to existing things or states of affairs. But if so, then The Little Leeway Thesis gives positive reason to doubt the metaphysical possibility of such things and states of affairs. And if that's right, then we have a defeater for any *prima facie* justification such thought experiments might have enjoyed. Below are two concrete examples in which our scientific knowledge conforms to, and thus supports, the Little Leeway Thesis, and thereby defeats the target possibility claims of exotic thought experiments:

---

[8] I'll offer a reasonably precise account of the distinction between "exotic" and "humdrum" thought experiments and possibility-claims later on in the chapter, when I sketch a positive account of our knowledge of possibility. For now, it's enough to speak in these rough and intuitive terms.

19

Case 1: Floating Iron Bar[9]

Changing the properties of a bar of iron so as to make it float requires changing its gravitational properties, which in turn requires changing its specific density, which in turn requires changing either the size of its unit cells or the number and kinds of atoms within them. But to change the latter requires a change in species, in which case it's no longer iron. And to change the former results in a change in the features determined by the lattice structure of its unit cells (cleavage, crystal habit, optical properties, etc.), in which case, again, we're no longer talking about iron. Either way, then, changing the properties of iron so as to make it float results in non-iron. In fact, iron is a transition metal, lying between manganese and cobalt on the Periodic Table. Thus, the range of possible densities for iron is highly restricted by these neighboring transition elements. But such leeway for changes in density isn't anywhere near what's required to make iron float. Thus, floating iron bars are metaphysically impossible.

Case 2: Carnivorous Rabbits[10]

Changing the properties of a rabbit so as to make it a carnivore would require changing (e.g.) its *dental* features (it lacks meat-tearing canines) and its *digestive* features (it's designed for rapid digestion of vegetation that results in small edible pellets for re-consumption -- digested meat from such a system would be toxic), etc. Further, The Principle of Natural Selection implies that organisms must be broadly adapted to their environmental niches. Now since carnivores are predators while rabbits are prey, a host of the rabbit's current features are useful only for making it adaptable, and thus viable, as a prey and not as a predator. Thus, it will have to be changed in these respects as well so as to make it adapted to its niche *qua* predator. So, for example, we will have to alter its scut tail, which functions as a signal to the group in contexts of danger, and thus not an adaptive predator trait, which often requires stealth. It would also require a radical lowering of the rabbit's birth rate, so as not to quickly deplete surrounding food resources. But at this point in our alteration of the rabbit, (but probably much before this point) it no longer seems to be a member of the class of lagomorphs. Indeed, such alterations in rabbits would require radical changes in the evolutionary history of the lagomorphs, etc., etc. But if these things are so (and there are plenty more where those came from), then to change a rabbit from an herbivore into a carnivore would require changing it into something other than a rabbit, indeed it would no longer be anything even remotely resembling a rabbit

---

[9] The following case is a paraphrase of Seddon's in *ibid.*, pp. 485-86.

[10] The following case is a paraphrase of Seddon's in *ibid.*, pp. 491-92.

(thus, even in the biological realm, a change in the "nodes" requires a change in the "root"). Thus, carnivorous rabbits are metaphysically impossible.

In short, our well-confirmed theories imply an interrelation among natural kinds, interdependence of the features of things within a given kind, and highly specific niches for each kind to occupy. And these facts give good evidence that there is little leeway for the possibility of a few features of a thing being altered without a change in thing or kind of thing – indeed, the resultant kind may be impossible, as there is good reason to think there is no possible niche for it to occupy in many, if not most, cases. But if so, then we have good reason to think that exotic thought experiments that imply otherwise are unreliable guides to possibility.

But not only do we have good evidence that particular theories imply the Little Leeway Thesis *within* those theories; we also have the basis for a *track-record* argument from the set of theories to the meta-claim that the Little Leeway Thesis applies to things *in general*. But exotic thought experiments represent things in ways that require the falsity of the Little Leeway Thesis. And if that's right, then any justification that exotic thought experiments might have conferred on possibility claims is defeated. We can put the problem the Little Leeway poses for exotic thought experiments in terms of the following argument:

> **The Argument From Little Leeway**:
> 1. The Little Leeway Thesis is true (supported by well-confirmed theories).
> 2. If the Little Leeway Thesis is true, then we have good reason to think the things represented by exotic thought experiments, though conceivable, are impossible.

21

3. If we have good reason to think the things represented by exotic thought experiments, though conceivable, are impossible, then exotic thought experiments are unreliable guides to possibility.

4. Therefore, exotic thought experiments are unreliable guides to possibility.

In short, scientific theories tell us that the structure of the world is tightly integrated, and in a way that implies the Little Leeway Thesis. But the legitimacy of exotic thought experiments as evidence of possibility requires the falsity of the Little Leeway Thesis. Therefore, we have reason to think that exotic thought experiments aren't good evidence of possibility.

1.3 "Not So Fast!" Roles, Occupants, and Epistemic Possibility

So far, we have looked at one consideration that suggests that, while me may know a good many possibility claims, there are tightly-circumscribed bounds to the scope of such knowledge. However, one might think there is an easy way to respond to this worry in a way that avoids the implication of mitigated modal skepticism. For it seems that we can screen off the false positives above by distinguishing between possibility claims about actual things and kinds, on the one hand, and possibility claims about non-actual kinds that play (or partially play) the *roles* of actual kinds, on the other. Thus, carnivorous rabbits are one thing, and carnivorous creatures that sort of *look* like rabbits are another. And while Seddon's argument may tell against the former possibility, they have no bearing against the latter.[11]

---

[11] David Chalmers has developed a rigorous and sophisticated account of modal epistemology that trades on this distinction between possibility-claims about roles vs.

I have serious doubts about whether such things are possible[12], but let that pass. For the worry now is how we could have knowledge, or even reasonable belief, that such things or kinds are metaphysically possible. Now at first, the answer seems trivially easy: we can conceive[13] that such things are possible. However, two things need to be asked about this answer. First, can we really conceive such things? And second, are the sorts of conceivings involved in such cases reliable – or at least, are they of a sort that can confer at least *prima facie* justification on such beliefs? These questions will receive more detailed discussion in due course, but for now I'd like to make some important preliminary points that raise worries for the current response to The Argument From Little Leeway. Thus, consider the claim that one can conceive the possibility of floating iron-bar-like-things. This claim had better not mean merely that we fail to see, or are unable to rule out, that such a thing is metaphysically impossible; that is, that floating

---

those about the actual occupants of those roles. We'll take a close look at his account of modal epistemology in chapter 3.

[12] For example, doesn't Seddon's argument give us reason to think that *anything* we would call a creature or animal must fall into a nested hierarchy of species, genus, a class, etc., and that it must be broadly adapted to its environment? But then how could any predator like a carnivorous rabbit-like creature also have the apparently superficial features of a rabbit -- such as a scut – and yet be evolutionarily viable *qua* predator? And while we're at it, what about floating iron-like bars? Even if there could be such things, they would have such a remote structural resemblance to iron bars that I doubt they would have much relevance to the sorts of points often made in thought experiments. I'm not sure what could be said in reply to such doubts, so I leave them aside, and just register my strong suspicions.

[13] Different philosophers have used 'conceive' to refer to different sorts of mental acts. We'll need to be more precise when we discuss specific accounts of modal epistemology in later chapters, but here I use 'conceive' as a catch-all term that refers to any mental act that philosophers have proposed as an epistemic basis for modal claims.

23

iron-bar-like-things are *epistemically* possible. For there are good reasons to think that such an unfettered approach to modalizing is not to be trusted. For one thing, such an account of conceivability makes metaphysical possibility a function of one's ignorance: the less you know, the more possibilities there are[14]. Thus, if I didn't know that water is H20, it would be metaphysically possible for water to exist without H20. But (assuming the necessity of identity), this is impossible.

This brings us to a second problem with such an approach to possibility-knowledge: it admits too many false positives. So, for example, I can't rule it out as impossible that Goldbach's Conjecture (GC) is false, even after reflection. If this is enough to make it reasonable to believe that GC is possibly false, then given that GC is necessarily true if true at all, it's thereby reasonable to infer that GC is *actually* false. Despite this result, I'm hesitant to submit my findings to the *Annals of Mathematics*.[15] That's bad enough, but we can keep going. Thus, I fail to see any incoherence in the possible *truth* of GC, in which case the current approach to modalizing would imply that this is good evidence of its possibility. But now we have an even worse problem. For when we combine this result with that of the previous, it now seems we have good evidence for both the possible truth and the possible falsehood of GC, in which case GC is a *contingent* truth. But again, GC is necessarily true if true at all; therefore, there is

---

[14] A number of authors have made this point. See, for example, Yablo, Stephen. "Is Conceivability a Guide to Possibility?", *Philosophy and Phenomenological Research* 53 (1993), pp. 7-8.

[15] This criticism is raised by, e.g., Yablo, *ibid.*, pp.8-9, and Van Cleve, James. "Conceivability and the Cartesian Argument for Dualism", *Pacific Philosophical Quarterly* 64, (1983), p. 37.

24

reason to worry that our unfettered, freewheeling approach to modalization is unworthy of our confidence.

But the problems for the unfettered approach of epistemic possibility as a guide to possibility aren't restricted to our judgments about GC. Thus, consider Alvin Plantinga's modal ontological argument[16]. The key premise is that, roughly, there is a possible world at which a necessarily existent individual exists. But given Axiom S5 – the one that says that what's possibly necessary is necessary *simpliciter* -- the truth of the key premise entails that Plantinga's god exists. Now even on reflection, I fail to find any incoherence in the concept of God as described by Plantinga. So if epistemic possibility justifies claims about metaphysical possibility, I should conclude that Plantinga's God exists. But, so far as I have been able to tell, there have been few takers on Plantinga's ontological argument. Again, that's bad enough. But the same goes for a maximally excellent being's possible non-existence. Thus, consider Peter van Inwagen's "knownos"[17]. A *knowno* is a being that knows there are no necessary beings. Now I fail to see any incoherence in the notion of a knowno. Alternatively, I see no incoherence in a purely physical, contingent, yet metaphysically independent or "free-standing" universe: there are the fundamental particles, and all else logically supervenes on that. So if epistemic possibility is a guide to possibility, it's also reasonable for me to believe that there are no necessary beings. Thus, the assumption that epistemic possibility is a guide to metaphysical possibility once again leads us into trouble. We could keep going if that

---

[16] *The Nature of Necessity* (Oxford: Oxford University Press, 1974), pp. 196-221.

[17] *Metaphysics*, 2nd ed. (Boulder, CO: Westview Press, 2002), pp. 107-108.

weren't enough ("I fail to see that my disembodied existence is impossible..."), but perhaps what we've said is sufficient to conclude that an unfettered, freewheeling approach to modalizing such as that found in the use of epistemic possibility will not do. And if that's right, then we don't yet have a principled answer to the worries for possibility-knowledge raised by Seddon.

1.4 A Call for Restraint: Liberals, Conservatives, and Modal Appearance

The lesson so far seems to be that any reliable, or at least responsible, approach to forming possibility beliefs requires some constraints. The hope, of course, is for a set of plausible constraints that meet the following conditions: (i) they aren't so *permissive* as to allow for lots of false positives, and (ii) they aren't so *restrictive* as to preclude the justification of the possibility-claim about twin water, and other exotic possibility-claims at the heart of many philosophical thought experiments. Well, what sorts of constraints are plausible? Stephen Yablo has suggested a reasonable minimum requirement: any guide to metaphysical possibility must at least represent its referents as *possible* (as opposed to merely *failing* to appear as *im*possible). Compare: No experience that fails to represent its referents as *actual* qualifies as a candidate guide to the actual world around you. This is at least one reason why imagination fails as such a candidate. Similarly, epistemic possibility fails as a candidate guide to possibility, as it does not represent its

26

referents as possible; rather, it merely *fails* to represent them as *impossible*. Let's follow

Yablo in calling this constraint, *the modal appearance test*.[18]

A dizzying array of accounts of the epistemology of possibility have been

advertised and endorsed by various philosophers, both past and present. We will not go

into all of them here.[19] However, we can use Yablo's modal appearance test to divide

accounts into two broad sorts, and to aid us in evaluating them. Thus, let's call an account

Conservative if it requires conceivings to pass the modal appearance test; call an account

Liberal if it lacks this requirement.

Like a number of other Conservatives, Yablo argues that modal appearances are

similar to perceptual appearances[20]. In particular he argues that the two are similar in at

---

[18] "Is Conceivability a Guide to Possibility?", *Philosophy and Phenomenological Research* 53 (1993), pp. 1-42, esp. pp. 4-7. Two other philosophers who hold to the modal appearance constraint are James Van Cleve and Paul Tidman. Van Cleve ("Conceivability and the Cartesian Argument for Dualism", *Pacific Philosophical Quarterly* 64 (1983), pp. 35-45) distinguishes between what he calls "weak conceivability" and "strong conceivability", where the possibility that *P* is *weakly conceivable* iff one fails to intuit that *P* is impossible, and the possibility of *P* is *strongly conceivable* iff one can intuit that *P* is possible. According to Van Cleve, only strong conceivability is a reliable guide to metaphysical possibility. Tidman adopts Van Cleve's strong/weak conceivability distinction in "Conceivability as a Test for Possibility", *American Philosophical Quarterly* 31 (1994), pp. 297-309.

[19] Although we will cover two prominent sorts of accounts in later chapters: Yablo's imaginability account (chapter 2), and Chalmers' two-dimensionalist account (chapter 3). I discuss a large variety of accounts in Appendix II. I explicate and evaluate Descartes's account in Appendix I.

[20] James Van Cleve (1983) uses perceptual language in his discussion of modal intuition. E.g., his remarks about such intuition as "a kind of intellectual *vision*" (p. 36), and as a way of "*seeing* that a proposition is possible" (p. 37) (italics added) suggests that he accepts the analogy between perceptual and modal seemings as well.

27

least three respects. First, both include *representational content*[21]: just as a perceptual

appearance represents *P* as being *true*, so a modal appearance represents *P* as being

possible. Second, both are *non-factive*[22]: just as it can appear to me that *P* is true when in

fact *P* is false, so it can appear to me that *P* is possible when in fact *P* is impossible.

Finally, both have *doxastic* and *epistemic force*[23]: just as a perceptual appearance that *P*

motivates or *moves* me to believe that *P* is true, and on the basis that *P* seems *prima facie*

justified to me, so a modal appearance that *P* is possible motivates or moves me to

believe that *P* is possible, and on the basis that P's possibility seems *prima facie* justified

to me.

In short, there is a worry for Liberal accounts about whether the *inability* to see

that *P* is *impossible* is sufficient evidence to think that *P* is *possible*. Conservative

accounts attempt to redress this worry by requiring that it must also *appear* to one that *P*

is possible. The notion of modal appearance here is supposed to be relevantly similar to

that of perceptual appearance. And the significance of this is supposed to be that our

---

[21] "Is Conceivability a Guide to Possibility?", pp. 5-7, esp. p. 6.

[22] *Ibid.*, p. 5, fn. 12.

[23] *Ibid.*, pp. 6-7, esp. p. 6. Yablo calls this aspect of a modal appearance, *epistemic appearance*. Paul Tidman seems to take the doxastic and epistemic force of a modal intuition as an indicator of possibility. Thus, in his "Conceivability as a Test For Possibility" (*American Philosophical Quarterly* 31:4 (Oct. 1994)), pp. 297-309), he says that, "When we claim to just see that something is possible...we find triggered within ourselves *a strong disposition toward accepting*...the possibility of what is being considered...There is *a strong felt tug* inclining us toward belief" (p. 308), emphases mine.

modal appearances share the coveted innocent-until-proven-guilty status of *prima facie* justification enjoyed by their perceptual analogues.

We've seen that the epistemic possibility account of knowledge of possibility is a Liberal account, as it lacks a modal appearance test. We've also seen that there is at least an initial basis for thinking that any such account is inadequate. However, do Conservative accounts fare any better in addressing the worries we've discussed so far? I'd like to suggest that, at least initially, the answer seems to be, "yes and no": Yes, in that there certainly are at least *some* possibility-claims that pass the modal appearance test. So, for example, the relatively humdrum possibility-candidates I mentioned earlier --- that my daughter's ball can get stuck on the roof; that my table can be moved about; that my car can be painted another color – don't merely fail to appear impossible, but positively appear possible. On the other hand, consider our example claim of an alien substance XYZ that perfectly plays the watery role. Does it seem or appear possible? If you're like me, it does not. Rather, it appears neither as possible nor as impossible; its modal status is left *undecided* for me. In this regard, my representation of the alien substance is like my representation of the existence of Plantinga's maximally excellent god, and unlike my representation of my car being a different color or my daughter's ball getting stuck on my roof.

1.5 Recap

Before going on, let's pause for a moment, take a deep breath, and recap the key conclusions so far. We've been investigating whether our knowledge of possibility is

significant in scope – significant enough to encompass possibilities involved in many contemporary thought experiments, such as the one about twin water. We started with the observation that we have strong possibility intuitions in humdrum cases, but then raised a significant worry, with the Little Leeway argument, for the reliability of such intuitions in more exotic cases.

Next we mentioned the response of distinguishing between roles and occupants, granting the problem with exotic modal claims about actual-world occupants of those roles, and arguing that our knowledge of the possibility of non-actual occupants of those roles survives the argument. The worry then became how, even if such exotic possibilities exist, we could have knowledge or reasonable beliefs about them.

The initial rejoinder was that we fail to see that such things are impossible, and that failure to see an impossibility – i.e., epistemic possibility -- is a guide to such possibilities. We then replied that there are significant worries for such an account: it makes metaphysical possibility a function of our ignorance; it admits of too many false positives; and it fails to meet what seems to be a minimal requirement on any epistemology of possibility: that a guide to possibility must at least represent its referents as possible.

This suggested a way to both avoid the problems caused by the over-permissiveness of epistemic possibility, and to ground the positive epistemic status of possibility claims. Unfortunately, we saw that adopting the modal appearance test comes at a high price, as it seems to imply that the claim about the possibility of twin water is unjustified.

30

This brings us to the present point in our inquiry. We seem to have surfaced a dilemma. For either we adopt a Conservative account of the epistemology of possibility, and thereby accept a modal appearance test, or we adopt a Liberal account, and thereby reject such a test. But adopting a Liberal account seems *too permissive*. For while it allows the twin water modal claim to count as justified, it allows too many other things to count as justified besides. On the other hand, a Conservative account is *too restrictive*. For while it rules out the problematic cases allowed in on Liberal accounts, it rules out twin water as well. Therefore, on either a Liberal or a Conservative account, the epistemic credentials of the twin water claim are sub-par. But of course, if twin water can't be justified, *a fortiori* are a host of other even more exotic possibility-claims in the literature unjustified: claims about the possibility of (e.g.) personal fission, maximally excellent beings, conscious states realized by the population of China, Swampman, zombies, disembodied souls, etc., etc.

It therefore seems that exotic possibility-claims (e.g., the twin water claim) do not appear possible, but rather undecidable, like the status of GC and the key premise in Plantinga's modal ontological argument. And while we will have to look at prominent Conservative accounts in detail in chapters 2 and 3 to make any confident judgments about this, our tentative hypothesis is that although we do know a lot of humdrum possibilities, modal claims similar to the one about twin water are probably beyond knowledge or reasonable belief.

1.6 First Steps Toward A Positive Account: Van Inwagen

We seem to have gone as far as we can go without talking about the details of specific accounts of our modal knowledge. And in light of the tentative conclusions reached so far, it makes sense to start with a fairly humble account: A Conservative account, in that it includes the modal appearance test, and one that acknowledges the existence of humdrum modal knowledge, but which captures the idea that such knowledge is significantly bounded. As it happens, there is an extant theory that accords fairly well with these desiderata: Peter van Inwagen's modal epistemology. In this section, then, we will take a look at his modal epistemology, and then evaluate its strengths and weaknesses. This will put us in a position to provide at least the rudiments of an adequate account of our knowledge of possibility.

Radical Skepticism vs. Remoteness Skepticism

Van Inwagen is a modal skeptic of a peculiar sort.[24] To see of just what sort he is, it will help if we first say a few words about the sort of skeptic he is not. A common form of skepticism starts with the idea that we can know nothing with respect to a certain class of beliefs, because the kind of evidence that issues from the source that individuates that class is *inherently* incapable of conferring sufficient justification. One notorious version of this form of skepticism identifies the class of *perceptual* beliefs -- beliefs about the external world based on perception -- as unknowable, due to the inherent evidential shortcomings of perceptual experience. Thus, I don't know if I have hands, or even

---

[24] The following discussion is based on Van Inwagen's paper, "Modal Epistemology", *Philosophical Studies* 92 (1998), pp. 67-84, reprinted in *Ontology, Identity, and Modality: Essays in Metaphysics* (Cambridge UP, 2001), ch. 10. All references below are to the latter version.

32

whether there are material objects at all, for my evidence for such beliefs – my perceptual experiences – just doesn't have what it takes to render external-world beliefs known or knowable. Call skepticism of this sort, *Radical Skepticism*.

By contrast, there is another sort of skepticism that doesn't write off the relevant class of beliefs due to general worries about its source or basis -- the source or basis in question may well be capable of yielding knowledge or justified belief. The problem is that the source's capacity to justify beliefs is *severely limited*; in fact, its justification-conferring ability is limited to beliefs involving the practical concerns of daily life. So according to this form of skepticism, we may well have lots of knowledge regarding matters "close" to the practical concerns of daily life. But the source's justificatory power drops precipitously when it comes to beliefs significantly remote -- and perhaps even modestly remote -- from the familiar realm of common experience. Van Inwagen gives a helpful illustration of this sort of skepticism in terms of our judgments of distance by means of unaided sight.[25] Such judgments are fairly accurate when it comes to short-range distance claims, such as "that bookcase is about 10 feet from my desk", "my house is about 50 yards from the intersection", and even "that mountain is about 25 miles away". But when it comes to sight-based distance judgments about objects of significant distance, such as claims about the distance of our moon from the Earth, our capacity for making accurate judgments drops precipitously. Call skepticism of this sort, *Remoteness Skepticism*.

---

[25] "Modal Epistemology", p. 246.

In light of the previous distinction, we may categorize Van Inwagen's modal skepticism as of the Remoteness sort. Thus, Van Inwagen does *not* think that our sources of modal justification are *inherently* incapable of giving us knowledge or justified belief about what's impossible, possible, or necessary. Indeed, he is at pains to point out that we have lots of modal knowledge about matters that figure prominently in the practical concerns of daily life.

However, when it comes to modal matters remote from these concerns, we're kidding ourselves if we think we can know such things, or even have reasonable beliefs about them.[26] Thus, I can know such humdrum modal claims as that my car is capable of being red, that my table doesn't have its current location of absolute necessity, and that there could've been fewer cats than there are. On the other hand, my powers of modalization are just too feeble for me to know such things as whether the property of maximal excellence is possibly instantiated or whether there's a possible world in which I exist and yet nothing material exists -- or even whether there could've been naturally purple cows; as such, they're remote from the practical concerns of daily life.

---

[26] *Ibid*, pp. 246-247: "My own view is that we often do know modal propositions, ones that are of use to us in everyday life and in science and even in philosophy, but do not and cannot know...modal propositions like ["It is possible for there to be a perfect being", "It is possible that I exist and nothing material exist", and "It is possible that there exist vast amounts of suffering for which there is no explanation"]. I have called this position "Modal skepticism". This name was perhaps ill-chosen, since...I think that we know a lot of modal propositions, and..."skeptic" suggests someone who contends that we know nothing or almost nothing...however...there has been another sort of skeptic: someone who contends that the world contains a great deal of institutionalized pretense to knowledge of remote matters concerning which knowledge is in fact not possible...It is in this sense of the word that I am a modal skeptic."

34

We may thus broadly characterize Van Inwagen's Remoteness Modal Skepticism in the following way. Say that a modal claim is *High* if it's remote from the practical concerns of daily life (e.g., that maximal greatness is possibly exemplified); *Low* if it's close to such concerns (e.g., that my car could've been painted red); and *Middling* if it's somewhere between High and Low in terms of remoteness (e.g., that naturally purple cows are possible). Then we can characterize Van Inwagen's modal skepticism broadly as the thesis that *the vast majority of High and Middling modal claims aren't knowable or capable of prima facie justification*. With this broad characterization of Van Inwagen's modal epistemology in mind, we turn now to take a closer look at it.

The Modal Knowledge We Do Have: Sorts and Sources

As mentioned above, Van Inwagen thinks we know the truth-values of many modal propositions, and beyond those we can know, there are many that enjoy various degrees of justification. Whence comes such knowledge and justification? Van Inwagen gives different answers, depending on the sort of modal claim in question.[27] On the one hand, there are lots of modal claims that most philosophers, Van Inwagen included, take to be non-mysterious and unproblematic. Thus, consider the following claims:

1. It's necessary that all bachelors are unmarried.

2. It's impossible that there is a barber who shaves all and only those men who don't shave themselves.

3. It's necessary that 2+3=5

---

[27] The following discussion of Van Inwagen's views about the sorts of modal knowledge we do have is based on *ibid.*, pp. 246-251.

35

(1) is known by reflection on the meanings of words; (2) is known by rules of deductive inference; and (3) is known by mathematical reasoning. Thus, for these sorts of modal claims, Van Inwagen sees no mystery as to their sources.[28] Although it's a bit inaccurate, let's refer to these sorts of modal claims as *analytic*. Thus, call a modal claim *analytic* if its truth-value is determined, if determined at all, by such things as concepts and logical or mathematical principles, and call a modal claim *non-analytic* if its truth-value cannot be so determined.

Further, we can *extend* our modal knowledge by combining our analytic knowledge with propositions known via observation. Thus, consider

4. Possibly, there are cows.

(4) isn't purely "analytic" in the sense that (1)-(3) are. For while (4) involves a deduction from actuality to possibility, it wouldn't be justified without perceptual knowledge of cows. Thus, we also have non-mysterious modal knowledge grounded in a mix of analytic and empirical sources.

So a good deal of our modal knowledge isn't mysterious. However, we also have modal knowledge whose sources are unknown. As one might have guessed, this sort of modal knowledge is non-analytic in character. All modal propositions in this category are of two sorts: claims about unactualized possibilities, and claims about non-analytic necessities.[29] Thus, consider

---

[28] *Ibid.* Of course, one could rightly point out that mysteries remain here, but Van Inwagen's point is that for a large class of modal claims to be individuated below, we lack even these *preliminary* sorts of answers regarding their sources of justification.

[29] *Ibid.*

36

5. I could've bought a PC instead of a Mac.

6. Necessarily, water is H2O.

On Van Inwagen's account of modal knowledge, (5) – a claim about the ability to do otherwise -- is a piece of "basic" modal knowledge, in the sense that one can know it, or at least be *prima facie* justified in believing it, without inferring it from other beliefs. Furthermore, the warrant-conferring basis of (5) is utterly unknown according to Van Inwagen.[30] By contrast, (6) isn't a simple case of basic modal knowledge. To see this, consider the following deduction:

5.1. It's impossible for something to be composed of different stuff.

5.2 Water is composed of H20.

6. Therefore, water is necessarily H20.

How do we come to know or justifiedly believe (5.1) and (5.2)? Well, (5.1) is a piece of basic non-analytic modal knowledge, which is equivalent to the claim that things have their fundamental composition of metaphysical necessity. And (5.2) is a fact about how the world is put together; we learn from scientists that water is composed of hydrogen hydroxide. This is of course a piece of empirical knowledge. We can thus put these claims together to deduce the Middling modal claim that water is necessarily H20. In this way, we can extend our modal knowledge and justified belief considerably.[31]

At this point, we have a complete list of the sorts of modal claims Van Inwagen takes to be justified:

---

[30] *Ibid.*

[31] *Ibid.*

i) The "analytic" cases, which consist of claims that can be determined to be necessarily true or false by means of reasoning with logical, semantical, and mathematical principles

ii) Inferences from actuality to possibility

iii) "Basic" non-analytic modal knowledge concerning practical matters of daily life

iv) Inferential non-analytic modal claims reasoned out from a combination of basic non-analytic modal knowledge and empirical facts about how the world is put together

By now, one might well be wondering whether Van Inwagen can meaningfully be considered a modal skeptic at all: aren't the classes of modal claims listed above sufficient to derive the truth-value of just about any modal claim we wish? Thus, consider modal claims of type (iii): the class of basic, non-analytic modal knowledge. We've seen, via (6) above, that Van Inwagen takes us to know that determinism is false, due to our basic modal knowledge that we could've done otherwise. But if he allows modal knowledge as substantial as *that* to count as basic, then isn't the sky the limit?

No. For recall our characterization of Van Inwagen's modal skepticism as a form of Remoteness Skepticism: we may well have plenty of modal knowledge, but it's restricted to matters close to the practical concerns of daily life. Well, according to Van Inwagen, our basic knowledge of our ability to do otherwise is a belief of just this sort. By contrast, there is a huge raft of philosophical claims -- claims involving, e.g., the possible exemplification of maximal greatness, of disembodied existence, personal fission, and cognitive systems composed of the population of China -- that are remote from such concerns; as such, and unlike the belief that we could've done otherwise, they are beyond the reach of justification. Therefore, Van Inwagen has a principled way to

38

differentiate properly basic from improperly basic modal beliefs, and he can do so in a way that entails a worrisome degree of skepticism about modal claims.

In light of these distinctions, we are now able to give a tolerable characterization of Van Inwagen's Remoteness Modal Skepticism: that *the vast majority of High and Middling non-analytic modal claims aren't knowable or capable of* prima facie *justification*. This, then, is Van Inwagen's modal skepticism.

1.7 Problems for Van Inwagen's Account

What to make of van Inwagen's modal epistemology? As a starting point for evaluation, recall that according to Van Inwagen, we have at least some non-analytic knowledge of various possibilities, and that such knowledge isn't based on other propositional evidence – at least not completely. Rather, such knowledge is basic, in the sense that it is non-inferential and yet *prima facie* justified. However, as we have seen, Van Inwagen has no positive account of the *source* of the justification for these beliefs. This is of course unsatisfying with respect to our current inquiry into the nature and scope of our knowledge of possibility. But more to the present point, it makes Van Inwagen's account subject to two related problems.

First, it makes the selectiveness of his modal skepticism seem *unprincipled and ad hoc*. For on the one hand, he wants to hold onto common sense, and say that we have at least some knowledge of possibility, viz., that of the humdrum sort. But on the other hand, he wants to say that the extent of such knowledge doesn't go beyond – or not much beyond – the humdrum. The former is basic, while the latter is not, and thus the latter

39

requires justification in terms of inference and argument. Now if he had a view about the source of justification for possibility-beliefs, then he could say that the former issue from that source, while the latter do not. And in this way, he could provide a principled basis for saying that some possibility-beliefs are justified, while others are not. However, Van Inwagen thinks we are completely *in the dark* as to the sources of justification for these beliefs.[32] Now one might think that an answer to this question is staring him in the face. For there are a plethora of accounts of the basis of our knowledge of possibility, and at least one of these may well be correct. However, for a number of reasons, van Inwagen finds all such accounts unsatisfactory.[33] Thus, without an account of the source of justification for our possibility-beliefs, his selective skepticism is unprincipled, and thus appears *ad hoc*.

This unprincipled selectiveness gives rise to the second problem with his view: it's *unstable*. To see this, consider the following commonly-assumed rationale for why conceivability must be at least *prima facie* evidence for possibility:

> **The Popular Argument**: If conceivability isn't at least *prima facie* evidence for possibility, then we aren't justified in believing even humdrum possibility claims (e.g., that my coffee table could've been two feet to the left of where it is now). But obviously, we *are* at least *prima*

---

[32] *Ibid.* p. 250: "Although I do not doubt that we have some modal knowledge, *I regard much of this knowledge as mysterious*. Some modal statements, I have said, we know by reasoning from what I have called "basic" modal knowledge – simple, obvious modal statements whose truth we are somehow in a position to know --... But how do we get started in this reasoning? How do we know the "simple, obvious" modal statements to be true? What is the ground of "basic" modal knowledge? *I do not know how to answer these questions.*" Italics mine.

[33] A systematic investigation and presentation of his reasons for rejecting such accounts falls outside the bounds of our current inquiry. However, I discuss his reasons in some detail in Appendix II.

40

facie justified in believing humdrum possibility claims; therefore, conceivability is at least *prima facie* evidence for possibility.

What I'm calling 'the Popular Argument' is analogous to a Moorean response to Radical Skepticism about perceptual knowledge. Thus, one might respond to the latter sort of Radical Skeptic by saying that if we're not entitled to trust our perceptual experiences as a basis for our beliefs about the external world without first justifying the reliability of perception, then I'm not entitled to hold such humdrum perceptual beliefs as that this is a hand. But obviously, I do know that this is a hand. Therefore, I am entitled to trust my perceptual experiences as a basic source of information about the external world. And while such experiences are not infallible, in that they do not entail the existence of the actual external world, they are nonetheless a source of defeasible, *prima facie* justification for such beliefs.

Similarly, the proponent of the Popular Argument responds to the Radical Modal Skeptic by saying that if he's not entitled to trust our conceivings as a basic source of information about what's possible, then he's not entitled to hold such humdrum possibility-claims as that his table can be moved about, that his car can be painted a different color, or that his daughter's ball could get stuck on the roof. But obviously, he does know such claims. Therefore, he's entitled to trust his conceivings as a basic source of information about what is metaphysically possible. And while such conceivings are not infallible, in that they do not entail that their putative referents are metaphysically possible, they are nonetheless a source of defeasible, *prima facie* justification for such beliefs.

The Popular Argument, then, when combined with van Inwagen's mysterianism about the source of our knowledge of possibility, poses a credible threat to the plausibility of Van Inwagen's Remoteness Modal Skepticism. For (i) if he admits that we have at least some non-analytic possibility-knowledge, (ii) if he has no account to take the place of the accounts of conceivability he rejects, and (iii) if The Popular Argument gives us reason to think that it *must* come from conceivability if we are to have any such knowledge at all, then there's pressure to reason, in G.E. Moore Shift fashion, that *something* must be wrong with Van Inwagen's case against conceivability as evidence of possibility, even if we can't say what that is. Therefore, without a positive account of the source of our knowledge of possibility, Van Inwagen's modal skepticism looks to be unstable.

In short, Van Inwagen holds to a form mysterianism about the source of our knowledge of non-analytic possibilities. This leaves him open to the charge of holding to an unprincipled selectiveness in his skepticism about certain sorts of possibility-claims. And this, in turn, raises serious worries about the stability of his position. Therefore, if our tentative hypothesis of Remoteness Modal Skepticism is to survive the criticisms that plague van Inwagen's view, we will have to come up with a positive account of the sources of justification for our possibility-beliefs.

We've seen that van Inwagen's mysterianism about the sources of possibility-knowledge undermines one's credence in his views about the scope and limits of such knowledge. However, despite Van Inwagen's disavowal of a specific account of our knowledge of possibility, some of his remarks are at least suggestive of a positive

account. It will prove instructive to briefly explore such an account, as it will shed some light on a basic source of possibility-knowledge.

To start off, van Inwagen asserts that we have knowledge of possibility-claims "that are *of use to us* in everyday life and in science and even in philosophy".[34] He thus seems to suggest here that a possibility-claim's *usefulness* is an indicator of whether we can know it.

In other places, his remarks suggest something slightly more specific. Thus, recall that we introduced the notion of Remoteness Modal Skepticism by means of van Inwagen's use of an analogy between modal knowledge and sight-based distance judgments. Given our current aim, it's worth quoting the relevant passage at some length:

> One way to get an intuitive grasp of what I mean when I speak of "modal skepticism" is to consider the analogy of distance. In my view, many of our modal judgments are analogous to judgments of distance made by the eye…There are, however, circumstances – circumstances remote from the practical business of everyday life – in which they are not accurate at all…. Analogously, I should say, we are able to discern the modal status of some propositions in a way that, like our intuitive judgments of distance, is "non-inferential".[35]

Thus, in this passage, van Inwagen suggests that knowledge of possibility and knowledge of distance are analogous, in that in both cases, we're reliable when the claims are close to practical concerns, but unreliable when the claims are remote from such concerns. Van Inwagen alludes to the same point when he says:

> …I am convinced that, whatever it is that enables us to determine the modal status of ordinary propositions about everyday matters, this method

---

[34] "Modal Epistemology", p. 245. Emphasis added

[35] *Ibid.*, p. 247. All italics but the last added.

43

or mechanism or system of intuitions or whatever it should be called is of no use at all in determining the modal status of propositions remote from the concerns of everyday life.[36]

We can use these passages to give an account of our knowledge of possibility. Where 'P' denotes a non-analytic possibility claim:

(PVI) S is in a position to know whether P is true iff P is close to the

practical concerns of S's daily life.

Thus, according to PVI, I'm in a position to know that I could've bought a PC instead of a Mac, since voluntary choices (at least of this sort) are close to the practical concerns of my daily life, and this puts me in a position to reliably judge of its possibility. However, I'm not in a position to know whether there could've been a naturally purple cow, as that is remote from the practical concerns of my daily life, and thus I'm in no position to assess whether naturally purple cows are possible. In short, according to PVI, a possibility-claim's figuring into one's practical concerns in a salient way somehow puts one in a position to know it.

What to make of PVI? Unfortunately, it appears that PVI is inadequate. First, closeness to one's practical concerns isn't necessary for knowledge of possibility. So, for example, I know that it's possible for the state capitol building in California to be painted orange, and yet this possibility is remote from my practical concerns.

Furthermore, closeness to one's practical concerns is not sufficient for knowledge of possibility. So, for example, the actions that flow from my considered desires and preferences are such that they're close to the practical concerns of my daily life, and yet

_____

[36] *Ibid.*, pp. 252-53.

I'm in no position to determine whether it's possible that I had done other than what I actually choose to do in those situations. In fact, there is considerable scientific evidence that I could not have done otherwise in those situations, despite my tendency to think the opposite. For these reasons, PVI is an inadequate account of justified possibility-beliefs.

However, despite the inadequacy of PVI, it appears to point in a promising direction. For although our knowledge of non-analytic possibility-claims isn't quite captured in terms of closeness to one's practical concerns, the account picks up on the broader idea that *there is a strong connection between knowledge of possibility and our actual experience.* I'll try to sketch a more accurate account of that connection in the next section.

1.8 Eliminating the Mystery: Toward a Positive Account of Sources

To start off, I want to briefly discuss a clue as to the basis of our knowledge of possibility. I will argue that the pattern of fluctuation in our confidence among a range of possibility-beliefs is evidence that our knowledge of possibility is grounded in our knowledge of the actual world. Thus, consider the following propositions:

1. My table can be three feet to the right of my chair.

2. There can be naturally purple cows.

3. There can be an alien substance that plays perfectly the watery role.

4. There can be minds without bodies.

5. There can be an Anselmian being.

Does your degree of confidence in the possibility claims listed progressively deflate as you go down the list? Mine, too. Is your degree of confidence in these claims enhanced by trying to conceive them? Mine, neither. But notice that as we go down the list, the referents of the modal statements progressively decrease in similarity with the actual world and with our knowledge of how the actual world works. Thus, there is a tight correlation between a possibility-candidate's degree of similarity to the world as we know it, on the one hand, and our degree of confidence in the truth of the candidate's corresponding modal statement, on the other.

What accounts for this? I want to suggest what I take to be a natural explanation: a given possibility statement is *justified* for a given person just to the extent that it depicts either a token of a type of something in the actual world, or of something relevantly similar to such a type or token. For suppose otherwise; suppose our knowledge of possibility were grounded in our conceivings irrespective of whether they depict, or are relevantly similar to, things in the world as we know it. Then we wouldn't expect there to be such a tight correlation between our degree of confidence in a given possibility claim and the extent to which it resembles, or is grounded in, our knowledge of the actual world. Rather, we would expect it to correlate with our ability to conceive the referent of a given possibility claim, *irrespective* of its degree of similarity to the actual world. But, at least initially, this doesn't seem to be so: conceiving the relevant entities makes little or no difference in our confidence in the corresponding possibility claim[37]. Therefore, at

_____

[37] Of course, one might worry that we're not conceiving the right sorts of entities, or that even if we are, we're not conceiving them in the right way. And of course, there have been a number of accounts, beginning at least with Descartes, of the proper entities to

46

least initially, it appears that the tight correlation between our degree of confidence about the truth of a possibility statement and its degree of similarity to the actual world provides at least some evidence that the latter is the source of justification for our beliefs about what is possible.

If the Confidence Argument is on track, then we have at least some reason to think that knowledge of possibility is grounded in our knowledge of the actual world. However, we don't yet have a clear idea of *the way(s) in which* the latter grounds the former. In what follows, therefore, we will look at a number of cases in which it's clear that we have knowledge, or at least justified belief, about the relevant possibility claim, with the hope of finding the ways in which they trace back to our knowledge of the actual world.

To start with the simplest sort of case, I know that

1. My car is painted gold.

is possible, since I know that it's *actual* – I've seen my car, and the color of its paint is gold. My belief that (1) is possible can therefore be supported via a direct inference from actuality to possibility. So here is a simple, but clear, limiting case of knowledge of possibility grounded in knowledge of the actual world.

Moving to only slightly less unremarkable cases, consider:

2. My desk is moved to the middle of my office.

3. My car is painted blue.

---

conceive, and of the right ways to do so. We will discuss the more prominent contemporary versions of such accounts in chapters 2 and 3; I discuss Descartes' account in Appendix I.

Like (1), only the most radical sort of modal skeptic would deny that we can know that (2) and (3) are metaphysically possible -- at least for anyone in my epistemic position. And like (1), belief that (2) and (3) are possibly true can be supported by our knowledge of the actual world. But unlike (1), neither (2) nor (3) is *actually* true. In what way, then, are they grounded in our knowledge of actuality? One way to account for such knowledge is via an inductive inference from actual to non-actual *tokens*. To be more specific, one may support their belief in the possible truth of (1) and (2) by reasoning as follows: based on perception, I know that other tables have been moved around in rooms, and that other cars have been painted different colors. Those states of affairs are therefore possible, since they're actual. From such observations, I reason inductively that since a number of tokens of those types of states of affairs are possible, the token states of affairs denoted by (2) and (3) are probably possible as well: it's probable that my office table can be moved to the center of my office, and that my car can be painted blue.

However, even if one *could* reason in the way sketched above, it seems much less natural and much more tentative than an alternative way of reasoning from types to tokens when it comes to cases like (1) and (2). Thus, one might instead justify the possibility of such claims as follows: we can conceive of various types of states of affairs. Some of these are possible and some are not. However, we don't know which ones are possible unless or until we observe some of their actual tokens. Actual tokens of a type of object or state of affairs thus function as demonstrations of the intrinsic metaphysical possibility of objects or states of affairs of that type – whether there is just one such token or infinitely many. Therefore, the possibility of conceived types of objects or states of

48

affairs can be justified when they are backed by observational knowledge of one or more actual tokens of them, via *rational induction.*

But what if I haven't observed the tokens of types of states of affairs relevant to assessing (2) and (3)? It seems that I may yet be able to have knowledge or justified belief about those statements. For someone else may *tell* me that they've seen such things. My modal knowledge would thus be derivable from the observations of others via *testimony.*

Now suppose I have neither observational nor testimonial knowledge of cars being painted new colors or of tables being moved around. Am I then unable to have knowledge, or justified belief, about the two claims? No. For if I've seen other sorts of objects moved around and painted, I can use these observations as a partial basis of a *folk theory* of how the world works. Our folk theory includes not only a folk physics about the behavior of inanimate objects, but also a folk psychology about the mental states and behavior of oneself and others.[38] The theory receives confirmation in the usual ways that scientific theories receive confirmation, such as its ability to explain and predict the world as we experience it. On the basis of our folk theory of how the world works, then, I can reason that since ours is the sort of world in which middle-sized solid objects can be

---

[38] Timothy Williamson has independently appealed to our folk physics and folk psychology as at least a partial basis of our knowledge of possibility. The primary difference between his view and mine on this point is that Williamson ties our modal knowledge to our facility with counterfactual reasoning, whereas I make no commitment to such a connection. For a thorough exposition and defense of Williamson's account of modal epistemology, see his *The Philosophy of Philosophy* (Oxford: Oxford University Press, 2007).

49

moved around and painted different colors, it's possible for my table to be moved to the center of my office, and for my car to be painted blue.

Suppose the sorts of cases discussed above largely exhaust our knowledge of possibility. Still, it seems that we nonetheless have some *justified beliefs* about various possibility claims that go, at least to some degree, beyond them. Thus, consider van Inwagen's naturally purple cow case.[39] Van Inwagen is pessimistic about having reasonable belief about the possibility of naturally purple cows. He asks us to consider the following modal proposition:

1. Possibly, there are naturally purple cows.

He then points out that:

> "A philosopher will confidently say that a (naturally) purple cow is possible. But he or she will not in fact have devoted any thought to the question whether there is a chemically possible purple pigment such that the coding for the structures that would be responsible for its production and its proper placement in a cow's coat could be coherently inserted into any DNA that was really cow DNA – or even "cow-like-thing-but-for-color" DNA...Either the structural formula for such a pigment is already there, lurking Platonically in the space of chemical possibility, or it is not. And – so far as I know – no one has any reason to assign any particular subjective probability, high or low or middling, to the thesis that it is lurking there."[40]

Van Inwagen thus concludes that the epistemic status of (1) is inscrutable. However, this seems to be too extreme. For, granted, we may not *know* whether (1) is possible, and for the reasons he mentions here. But it's hard to agree with him that (1) is *inscrutable*. For while it's true that there are no naturally purple cows, there are nonetheless *cows*, and we

---

[39] "Modal Epistemology", p. 254.

[40] *Ibid.*

50

know that these are naturally of various colors – though not, of course, purple. But if so, then there is good reason to be confident that the type of state of affairs involving the existence of naturally purple cows is *relevantly similar* to the actual states of affairs involving cows of various colors that we find in the actual world. And given that arguments from analogy are legitimate, we have reason to think that since the latter are possible (because they're *actual*), so, probably, are the former. In this way, (1) receives at least *some* epistemic support from an argument from analogy between states of affairs involving actual cows and those involving naturally purple cows. Of course, it is no doubt often a messy and difficult task to determine whether the analogues in such arguments are sufficiently similar, and in the relevant observed respects, to justifiedly conclude that they are probably similar in the unobserved respect of being possible. However, this is true of arguments from analogy *in general*, and thus poses no *special* problem for arguments from analogy when applied to possibility claims.

We've seen that our knowledge of possibility can be traced back to our knowledge of actuality in various ways: to deductions from observations of actuality; to inductive, track-record evidence from tokens of the same types of observed actualities; to rational induction of single tokens of a given type of observed entity; to our folk theories of how the actual world works; to arguments from relevant similarity with the actual world; and to testimony of all the above. So far, though, we've only used these methods to justify relatively uninteresting possibility claims. However, as I will now argue, such methods can be extended to a number of philosophically interesting cases as well. So, for example, consider the claim that

51

4. There is justified true belief without knowledge.

Can the possibility of (4) be supported solely by information about how the actual world works? I think it's clear that it can. To see this, consider the following variation on Keith Lehrer's Nogot/Havit case[41]: Nogot, one of your colleagues, doesn't own a Ford, but you have excellent but misleading evidence that he does: he's long expressed the wish to own one, and you've seen him drive one to work for the last several weeks; you've seen something that looks just like a registration form with his name on it, etc. From this evidence, you justifiedly form the false belief that Nogot owns a Ford. And from the latter belief, you validly infer that someone at your work owns a Ford. As it happens, but unbeknownst to you, your colleague, Havit, owns a Ford. So your deduced belief is true and justified, but not knowledge.

The Nogot/Havit case is clearly different from the last two. In particular, it involves certain sorts of elaborate deception that few have experienced in the actual world. Nevertheless, our knowledge of its possibility can be supported by what we know of the actual world, and for reasons similar to those discussed in the previous cases. Thus, we have experience of people owning things, including cars; we have experience of ourselves having evidence for a false belief, as well as the testimony of others having such evidence; we have experience of people intentionally tricking others with false or misleading evidence; we also have experience and testimony of cases where people infer true things from false things. Finally, we have knowledge of our own psychology that we sometimes believe misleading evidence, and make fallacious inferences. We can thus use

---

[41] Lehrer, Keith. "Knowledge, Truth, and Evidence", *Analysis* 25:5 (1965), pp. 168-175.

52

our observations and the testimony of others as the partial basis of a folk theory that justifies our belief in the possibility of the Nogot/Havit case.

It appears, then, that our account can explain our knowledge of possibility for at least one philosophically interesting case. But a little reflection will allow us to see the potential of our account to explain our knowledge of possibility in a number of other philosophically interesting cases, such as: the trolley problem case in the ethics literature; Locke's locked room case and Frankfurt's "Black and Jones" counterexamples against PAP in the free will literature; the fake barn case in the epistemology literature; the ship of Theseus case in the personal identity literature, etc. Thus, these cases, and many others, can be justified by our knowledge of how the world works, where this is understood in the ways sketched above.

To recap: the Confidence Argument suggests that our knowledge of possibility is grounded in our knowledge of the actual world. Our knowledge of possibility traces back to our knowledge of actuality in various ways: to (i) deductions from observed or testified-to actualities; (ii) inductive, track-record evidence from tokens of a type of observed or testified-to actuality (i.e., token A of type F is possible (since actual), and so other tokens of F are probably possible); (iii) inferences via rational or intuitive induction from single tokens of observed actualities to all tokens of that type; (iv) our folk theories of how the actual world works; and (v) arguments from relevant similarity with the actual world. In these sorts of ways, we can ground our knowledge and justified beliefs about what is metaphysically possible. Finally, we've seen that this account of our knowledge of possibility is sufficient to justify a number of philosophically interesting thought

53

experiments, in which case the latter can be properly used to evaluate their corresponding philosophical claims.

We have sketched an account of modal knowledge and justified belief. It builds off of the basic insights of van Inwagen's work on the topic, but attempts to advance it, at least to some degree. Thus, like van Inwagen's account, my account requires that possibility claims must satisfy the modal appearance test; both are therefore Conservative accounts of modal knowledge. Also, like his account, mine argues that our knowledge of non-analytic possibility is limited, for the most part, to the relatively humdrum. It therefore occupies the same genus with his account, as both are species of Remoteness Modal Skepticism.

On the other hand, our account of the epistemology of possibility is an advance over van Inwagen's in a number of ways. First, ours gives *a positive account* of the sources of justification for our possibility-beliefs. As we have seen, van Inwagen offers no such account. We've also seen that this is a significant deficiency in his account. For without such an account, it's hard to see how he can explain why we know some possibility claims and not others. But if not, then his selective modal skepticism is open to the charge of being arbitrary and *ad hoc*. And the lack of a principled, explanatory basis for distinguishing between modal knowledge and modal ignorance, when coupled with the Popular Argument, threatens to undermine the plausibility of his position.

By contrast, with our positive account of knowledge of possibility as grounded in our knowledge of actuality, we can at once explain the possibility-knowledge we do have, and offer a principled basis for distinguishing it from that which we don't. In fact,

our account provides a way to make the High/Low/Middling distinctions much more precisely than on van Inwagen's account, and in a way that *explains* why Low and Middling possibility claims are justified, while High possibility claims are not. Thus, let's distinguish between three sorts of possibility claim as follows. First, say that a possibility claim is *Low* iff it's grounded in our knowledge of the actual world in the ways mentioned in (i)-(iv). Second, say that a possibility claim is *Middling* iff it's not a Low possibility claim, but it's nonetheless grounded in our knowledge of the actual world in the way mentioned in (v). And finally, say that a possibility claim is *High* iff it is neither a Low nor a Middling possibility claim – i.e., it can be properly grounded in none of ways (i)-(v) listed above. Then we can say that a possibility claim is justified iff it's a Low or Middling possibility claim.

Unfortunately, possibility claims about alien stuffs playing the roles of actual kinds can be justified by none of ways (i)-(v), and thus are High possibility claims; they are therefore unjustified. In this way, we can give a principled basis for accepting humdrum possibility claims, while not accepting the exotic claims. And if so, then we can fend off the worries of a selective and unprincipled skepticism that plague van Inwagen's mysterian version of Remoteness Modal Skepticism, and thus resist the force of the Popular Argument.

The previous point leads us to a final respect in which our account differs from van Inwagen's. For recall that on his account, our knowledge of non-analytic possibilities is basic, in that it's both *prima facie* justified and non-inferential. However, on our

55

account, such *prima facie* justification is *derivative*, in that it has its source in ordinary perception.

1.9 Back to Twin Earth

An unfortunate implication of our account is that we have a lot less modal knowledge than philosophers have commonly thought. This brings us back to Putnam's Twin Earth thought experiment. Recall that, at least at first appearance, it relies on the modal claim that, possibly, there is an alien substance that perfectly plays the watery role. Unfortunately, we've never observed such twin water; nor have we received testimony as to the existence of such a substance (residents of Twin Earth have yet to visit us to testify of its existence). We therefore cannot generate an inductive argument from actual to non-actual tokens; nor can we reason to its possibility via rational induction. Furthermore, our folk theory of the actual world cannot support the possibility of twin water. And finally, we have no reason to think that twin water is similar to actual water in the relevant respects to justify its possible existence via an argument from analogy. Therefore, if our account of our knowledge of possibility is at all on track, belief in the possibility of twin water is unjustified. *A fortiori* are we unable to justify that large host of possibility claims involved in various other thought experiments much more bizarre than Putnam's Twin Earth thought experiment: the possibility of zombies and disembodied souls in the philosophy of mind literature; fission thought experiments in the personal identity literature; Thomson's people seeds thought experiment in the abortion literature; Davidson's Swampman thought experiment in the philosophy of mind and language

56

literature; etc. Thus, our account offers little encouragement that such thought experiments support their corresponding possibility claims.

I submit, therefore, that my original claim so far appears to be vindicated: we don't know if twin water is metaphysically possible. But more generally, we seem to have at least some initial reason to think that our knowledge of possibility is much more limited than traditional Armchair Philosophers might have hoped. This is of course relevant to the primary aims of this dissertation. For recall the objection to Armchair Philosophy we encountered in the introduction:

> **The Basic Argument Against Armchair Philosophy (BAAAP)**
> 1. Armchair philosophy is legitimate only if thought experiments give us insight into the nature of things.
> 2. Thought experiments give us insight into the nature of things only if they justify modal claims.
> 3. Thought experiments don't justify modal claims.
> 4. So, thought experiments don't give us insight into the nature of things.
> 5. So, armchair philosophy is not legitimate.

Now if the arguments of the present chapter are on track, then this considerably dampens the prospects for defending Armchair Philosophy via a principled, unqualified rejection of (3). Thus, assuming the truth of The Master Claim (viz., that thought experiments tell us about the world), our preliminary investigations here suggest that they often don't do so by revealing *modal* facts about it. And to the extent that this is so, there is pressure to look for another connection between thought experiments and the world -- one that's fundamentally *non-modal* in character.

However, it would be premature to make any firm conclusions about such matters at this stage. For several sophisticated accounts of modal epistemology have been developed, and at least one of these might provide a plausible way to justify the modal

claim about twin water, as well as other High modal claims involved in a host of

philosophical thought experiments. We will therefore take a careful look at a

representative sampling of these in the next two chapters.

Chapter 2:
A Critique of Yablo's Modal Epistemology

In chapter 1, I offered some initial considerations in favor of Remoteness Modal

Skepticism. At the end of that chapter, we saw that skepticism of this sort raises a *prima*

facie worry for The Modal Epistemology Response to BAAAP. For if thought

experiments can only achieve this aim by means of revealing *modal* facts, and our

knowledge of modal facts is limited to those of the Low and Middling variety, then our

knowledge of the world via thought experiments is limited accordingly. Thus, assuming

the truth of The Master Claim, there is some pressure to look for an alternative, non-

modal connection between thought experiments and the world.

Perhaps, though, there is yet a way to justify High possibility-claims. For recall

that in chapter 1, I gave a sketch of an observation-based account of our knowledge of

possibility. Now, admittedly, that account takes us no farther than Low and Middling

claims. But maybe there are *other* sources of modal justification, and these provide a way

to justify High possibility-claims, such as the one about twin water we've been

discussing.

Stephen Yablo has developed an account of modal epistemology that holds out

hope of meeting these desiderata.[42] For on his account, there is a source of modal

justification that's independent of observation, viz., *imaginings* of a certain sort (to be

---

[42] Yablo, Stephen. "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research* 53 (1993), pp. 1-42.

spelled out below). In this chapter, therefore, we'll investigate whether Yablo's account can justify High possibility-claims. Toward that end, I will explicate and illustrate the basic contours of his account. Then, I'll offer two basic criticisms of it: (i) that it fails to justify High claims, in which case it's equivalent to my account in its results at best, and (ii) that since this is so, his core thesis – that imagination provides an independent source of modal justification – is undercut.

2.1 Yablo's Imagination-Based Account

As I mentioned above, Stephen Yablo's account argues that our knowledge of possibility is grounded in imaginings of a certain sort. We'll discuss some of the particulars of Yablo's account in due course, but for now, it's enough to speak in terms of the broad contours of his account. Thus, the basic idea is that possibility-claims are *prima facie* justified in virtue of imagining possible worlds[43] at which their de-modalized counterparts are true.[44] So, for example, the claim that, possibly, a ball is stuck on a roof is justified if I can imagine a world at which 'a ball is stuck on a roof' is true.

---

[43] Yablo admits that, strictly speaking, possible worlds are too large and complex for humans to imagine. However, he thinks it sufficient to imagine the salient *segment* or *portion* of a world when evaluating possibility-claims. He therefore speaks in terms of imagining *scenarios* – portions or segments of worlds -- as the basis of justifying modal claims. However, as ignoring this technical aspect of his account will not affect the points we will discuss, I will ignore it, and continue to speak in terms of worlds. For Yablo's remarks on these points, see *ibid.*, pp. 28-29.

[44] Cf. *ibid.*, p. 31: "...the task of conceiving p divides into two sub-tasks: imagining a possible world, and satisfying oneself that p is true in it."

60

We can think of Yablo's account as a "modal telescope" account: one's faculty of imagination is like a telescope that can look into other possible worlds; imagining a world is like looking at another possible world through the modal telescope; and if you're able to imagine a world that would make the de-modalized target claim true in it, then you have, in effect, spied a possible world through the modal telescope that vindicates the target possibility claim.

It's important to be clear about the sort of imagining Yablo has in mind. As the telescope metaphor suggests, the relevant sort of imagining on his account is not that of entertaining a coherent proposition or narrative – what Yablo calls *propositional* imagining. Rather, it is (roughly) that of having phenomenal imagery of a distribution of objects and properties – what Yablo calls *objectual* imagining.[45] There is thus a world of difference between understanding the proposition that, possibly, a pit bull bites my arm, and imagining a *pit bull* biting my *arm*. And imaginings of the latter sort are what Yablo's account requires for justifying possibility-claims.

A worry: the Unexcluded Alternatives Problem

Now one might worry that Yablo has made it too easy for possibility claims to count as *prima facie* justified. For consider the claim that, possibly, there is transparent iron. Now I can objectually imagine (e.g.) a scientist accepting the Nobel Prize before a cheering audience, who then thanks all those who helped him in his long and difficult journey to create transparent iron, and then holds up something that looks like a sheet of glass.[46]

---

[45] *Ibid.*, p. 28.

[46] *Ibid.*, pp. 31-32.

61

But, intuitively, that wouldn't *prima facie* justify the claim that transparent iron is metaphysically possible. For the imagining is *equally* supportive of other claims that are compatible with the *impossibility* of transparent iron. For example, the imagining equally supports the claim that, possibly, (despite the impossibility of transparent iron) a group of jokester scientists get together and fool the public that they've created transparent iron, using a piece of glass as their object of deception.[47] And since my objectual imagining can't rule out this alternative that's compatible with its falsity, it seems wrong to say that the imagining justifies the claim. Call this *the Unexcluded Alternatives Problem.*

"No worries": excluding alternatives with high standards

Yablo's account does not fall prey to the Unexcluded Alternatives Problem. To see why not, we'll need to look at a couple of details in his account. So far, we've said that Yablo's account of verifying a possibility-claim involves imagining a world at which the de-modalized target claim is true, where imagining a world involves imagining, roughly, a distribution of objects and properties. However, not just *any* such imagining counts as verifying a modal claim. Rather, such imaginings must meet two conditions: (i) the imagined scenario must appear to one as metaphysically possible (as opposed to merely failing to appear to one as impossible), and (ii) it must appear to one that the relevant de-modalized claim is true in that scenario (as opposed to merely failing to appear false).[48]

To get a better grasp of Yablo's two conditions of verification for possibility-claims, let's look at some concrete examples of success and failure with respect to

---

[47] The example is van Inwagen's. See his "Modal Epistemology", pp. 255-57.

[48] *Ibid.*, p. 31.

verification. First, consider the claim that, possibly, the thing that looks just like a flower before me sprouts new petals.[49] Now I can imagine a world in which this appears to me as true. Unfortunately, that world doesn't appear to me as possible; rather, it neither appears as possible nor as impossible – its modal status is left undecided. For as it turns out, the object it sits next to it is perceptually indiscriminable from it, and while I don't know which one is which, I do know that one of the two is a mere *flower-facsimile*. And given that real flowers can sprout petals while facsimiles cannot, I'm not at all confident that what I imagine is metaphysically possible. Thus, while it passes Yablo's second condition of verifying a possibility-claim, it fails the first condition.

Next, consider the claim that, possibly, Goldbach's Conjecture is false.[50] Now I can imagine a world in which I find an article in a leading mathematical journal with the title "A Refutation of Goldbach's Conjecture". This world appears to me as metaphysically possible. However, absent imagining the proof itself within the pages of the article, this imagined world is compatible with the *truth* of Goldbach's Conjecture. For my imagining leaves it open that (e.g.) the attempted proof contains a subtle error within it. So while the case passes Yablo's first condition of verifying a possibility-claim, it fails the second.

Finally, consider the claim that, possibly, a ball is stuck on a roof. I can imagine a world at which there is a ball lying against the upper side of a chimney on a slanted roof. In this case, the world imagined doesn't just fail to appear to me as impossible; it

---

[49] *Ibid.*, p. 32.

[50] *Ibid.*, pp. 31-32.

positively appears to me as metaphysically possible. Furthermore, the imagined world doesn't leave it ambiguous or undecided that a ball is stuck on the roof; rather, it positively appears to me as one in which a ball *is* stuck on a roof – I imagine it there, wedged behind the chimney. This case thus satisfies both of Yablo's conditions of verifying a possibility-claim; the claim is therefore *prima facie* justified for me.

We've seen that Yablo's imagination-based account of modal epistemology doesn't allow just *any* old objectual imagining to count as verifying a possibility-claim; rather, imaginings that are justification-conferring generate two sorts of seemings or appearances in the imaginer: (i) the appearance that the world imagined is metaphysically possible, and (ii) the appearance that the de-modalized target claim is true in that world. But if so, then Yablo's account steers clear of the Unexcluded Alternatives Problem. Thus, go back to the case of the transparent sheet of iron. Now, admittedly, we were able to imagine a world that appeared metaphysically possible to us, viz., one in which a scientist seems to be giving an acceptance speech for the Nobel Prize, and then holding up something that looks like a sheet of glass. So the imagined world passes the first of Yablo's conditions for justifying a possibility claim. However, it fails to meet the second condition. For the imagined world does not make the de-modalized target claim – i.e., that there is transparent iron -- appear true to one in it. Rather, it merely fails to appear false to one in that world. For as we saw above, the explicitly imagined portion of that world is compatible with the *non-existence* of transparent iron in that world. So, for example, for all we have imagined, no one has created transparent iron, the award ceremony was a hoax, and the sheet held up by the scientist is a sheet of glass.

64

Now of course, one could just *stipulate* that the ceremony isn't a hoax, but rather a *real* Nobel Prize ceremony, and that the transparent sheet held up by the scientist is a sheet of *iron*, and not glass. But now a new problem arises. For while it may be that the imagined world is one in which there is transparent iron in it, that world no longer appears to one as metaphysically possible; rather, it merely fails to appear to one as impossible.

To sum up: Yablo's account of justified possibility-beliefs is what we're calling a "modal telescope" account, since its core idea is that modal claims are justifiable in virtue of imagining possible worlds that make them appear true to one. This amounts to imagining a world that verifies the target possibility claim, where Yablo's account of verification involves two sorts of seemings or appearances: the appearance that the world imagined is metaphysically possible, and the appearance that the de-modalized claim is true in it. And the requirement of these two sorts of appearances in justification-conferring imaginings enables Yablo's account to get around the Unexcluded Alternatives Problem. We therefore have a *prima facie* plausible account of the epistemology of possibility.

2.2 Evaluating Yablo's Account

Even if correct, it can't justify High claims

What to make of Yablo's account? Although it makes a real advance over the simple epistemic possibility account in the ways we've discussed, there are at least two reasons for thinking it makes no substantive advance over our observation-based account. For,

first, suppose we grant, at least for the sake of argument, that Yablo's account specifies a trustworthy guide to metaphysical possibility, providing a basis for *prima facie* justified possibility-beliefs. The problem is that while the stringent constraints on his Conservative account screen off false positives, they also screen off *High* possibility claims, such as the twin water claim.

To see this, consider again the claim that, possibly, there is an alien substance that perfectly plays the watery role. Now to evaluate this claim via Yablo's method, we need to objectually imagine a world that meets the following two conditions: (i) it appears to us as metaphysically possible, and (ii) it appears to us that 'there is an alien substance that perfectly plays the watery role' is true in it. Unfortunately, while I'm able to objectually imagine a world in a way that meets either one of Yablo's two conditions, I'm unable to objectually imagine a world in a way that simultaneously meets both.

Thus, suppose I objectually imagine a world W in which I'm standing in front of a swimming pool full of gently undulating clear liquid on Twin Earth. Suppose further that I imagine a chemist in a lab coat in W pointing to the pool and saying, "Don't be fooled: that watery stuff is XYZ, not H20". Would *that* count as verifying the possibility of twin water? Pretty clearly not. For what I imagine of W is compatible with the claim that the chemist *mistakenly* or *deceptively* refers to the liquid in the pool as XYZ. Therefore, while W appears to me as metaphysically possible, and thus passes on (i), the de-modalized target claim, viz., 'there is an alien substance that perfectly plays the watery role', neither appears to me as true nor as false in W, but rather as undecided. My objectual imagining thus fails on (ii).

66

Now you might say that the only problem with the imagined world is the bit about the chemist. For if I take him out of the scenario, I remove the dubious intermediary between my judgment and the objectual representation of XYZ. However, this merely replaces one dubious judge for another. For consider a world W' just like W, but with the chemist removed. So now it's just the pool of undulating clear liquid and I. How am I to decide, just by objectually imagining the liquid in the pool, whether it's XYZ? For the objectual content of the scenario regarding what's in the pool does not represent itself to me as XYZ. Rather, it merely represents itself to me as *clear (undulating) liquid.* But if so, then since this aspect of the imagined world is *all I have to go on* in determining whether the stuff in the pool is XYZ (as opposed to, e.g., ordinary H20), my labeling it as XYZ is nothing more than a shot in the dark.[51] Thus, my competence in correctly labeling the liquid is no more to be trusted than the word of the imagined chemist. Therefore, as with W, my objectual imagining of W' passes on (i), but fails on (ii).

On the other hand, suppose we try to cut to the chase and just *stipulate* that the stuff in the pool in W' is XYZ, and not water. Given this stipulation, satisfying (ii) is trivial. However, the problem now is that I can't objectually imagine W' in a way that satisfies (i). For in this case, the non-XYZ watery stuff in W' appears to me as neither possible nor as impossible, but rather leaves its modal status *undecided.*

Our discussion above suggests a moral about Yablo's account: if it's correct, it seems to have the consequence that, short of the testimony of alien visitors or divine

---

[51] My worries about the stipulative contents of imaginings not backed by independent evidence have been deeply influenced by Peter Kung's work. For a penetrating and detailed discussion of this sort of problem, see his *Imagination and Modal Epistemology*, Ph.D. dissertation, New York University, 2004.

revelation, we can't know whether twin water is metaphysically possible. For it seems that no matter how I vary the scenario, one of two things will happen: either what I objectually imagine is compatible with the falsity of the claim about twin water, or it contains descriptions of things that I can't objectually imagine – at least not in a way that passes the modal appearance test. Thus, even if Yablo's epistemology of possibility is correct, it can't be used to justify the twin water possibility-claim.

But more significantly, the problem of justifying the twin water possibility claim on Yablo's account seems to generalize to *all* High possibility claims. For all such claims involve elements that go beyond what can be properly based on observation. They will therefore require justification in terms of objectual imagining, if his account is correct. But then the worry is that all such imaginings fall prey to the dilemma that applies to twin water: either we allow stipulations to count as part of the imaginings or we don't. If we do, then the stipulations will outstrip the content of what one strictly objectually imagines in a given world, thus undermining the prospects of an appearance of possibility. High imaginings that allow stipulations will therefore fail condition (i) of Yablo's account. On the other hand, if we don't allow stipulations to count as part of justification-conferring imaginings, then what we imagine will be compatible with the falsity of the de-modalized target claim in that world, and will therefore fail condition (ii) of Yablo's account. Therefore, it seems that even if Yablo's account of possibility-knowledge is correct, only Low and Middling possibility claims can be known or *prima facie* justified on his account, in which case it's equivalent to our account in its consequences.

68

Now one might argue that my doubts about the possibilities here are based on a form of Radical Modal Skepticism, and are therefore unwarranted. Thus, one might appeal to the Popular Argument in reply: if Yablo-style imaginings aren't a source of *prima facie* justification for possibility-beliefs, then I can't justifiedly believe that, for example, my daughter's ball can get stuck on my roof. But I can justifiedly believe that my daughter's ball can get stuck on my roof; therefore, Yablo-style imaginings are a source of *prima facie* justification for possibility-beliefs.

However, this response fails. For as we saw in chapter 1, we have an observation-based account of the epistemology of possibility, and this account grounds the *prima facie* justification of Low and Middling possibility-claims, while leaving High possibility-claims unjustified. But if so, then we have a principled basis for accepting the former sorts, while rejecting, or at least suspending judgment on, the latter sort. The key premise in the Popular argument is therefore defeated, and with it, the main reply to our criticism of Yablo's account.

Yablo's account is unmotivated

Second, and relatedly, I think we have reason to worry that Yablo's core idea – i.e., that imaginings provide *independent* evidence for possibility – is undercut. For his primary argument for his core idea is that it's the best explanation of the epistemic force of paradigm-case thought experiments. Thus, in considering whether we come to enjoy the appearance of some x's possibility in virtue of an objectual imagining of the sort he

prescribes, he asserts that this is the usual way that we come to learn of x's possibility, and then appeals to the Gettier cases and Putnam's Twin Earth cases as examples.[52]

But this line of reasoning in support of his core idea fails. For as we saw in chapter 1, we can account for the epistemic force of the Gettier counter-examples by means of our account of the epistemology of possibility, which explains our possibility-knowledge in terms of observation and observation-sensitive folk theory of the actual world.

What about Putnam's Twin Earth case? It's worth pausing to note how he refers to the case. For he makes no reference to the most exotic element of the thought experiment -- twin water --, but rather to the much more metaphysically and epistemologically unproblematic aspect of the case – the part about people just like us who refer to different things with their words.[53] Could it be that his avoidance of explicit twin water-talk is intentional? Whether it is or not, we saw in our discussion of the first criticism that Yablo's *own account* can't justify the possibility claim about twin water. But if so, then the twin water case doesn't support his core thesis, either.

To sum up: Yablo's account makes no advance over my account. This is for at least two reasons. First, even if his account truly captures a way to gain justified possibility-beliefs, it can't justify *High* possibility claims, such as the twin water claim. And second, his core thesis – i.e., that objectual imaginings provide independent *prima facie* justification for possibility claims -- is undercut. For he appeals to such mental

---

[52] *Ibid.*, p. 30.

[53] *Ibid.*

episodes as the best explanation of the epistemic force of paradigm thought experiments, such as the Gettier cases and the Twin Earth case. But the epistemic force of the former cases can be explained in terms of our observation-based account, and thus without appeal to anything beyond our knowledge of the actual world. And the epistemic force of the latter case cannot be handled at all on Yablo's *own* account. But if these things are so, then his abductive argument for imagination as an independent source of *prima facie* justification for possibility-beliefs fails.

I conclude by briefly returning to the larger aims of our inquiry. In this chapter, we continued our investigation into the nature and scope of our knowledge of possibility. But as in chapter 1, we failed to find a solid rational basis for *High* possibility-claims. The result is that an unqualified version of the Modal Epistemology Response to BAAAP is starting to look unpromising. However, it's still a bit too early to draw any firm conclusions about this, as there is another sophisticated account of modal knowledge to consider in the next chapter. However, if that account likewise fails to justify High possibility-claims, then given our assumption of the Master Claim, we'll have to look for another connection between thought experiments and the world besides the standardly-accepted *modal* connection.

Chapter 3:
A Critique of Chalmers' Modal Epistemology

3.1 Introduction

We've seen that when it comes to the scope of justified possibility-claims, Yablo's imagination-based modal epistemology takes us no farther than our observation-based account. For when it comes to High possibility claims, his account faces a dilemma: Either we allow mere stipulations into the content of imaginings or we don't. If we do, we undermine the imagined world's appearance of possibility; if we don't, we undermine the target claim's appearance of truth in that world.

However, despite the problems we've raised for Yablo's account, some might remain persuaded that imagination is good evidence for High possibilities. Thus, one might argue that from the fact that a particular *version* of imagination-based modal epistemology fails to justify High possibilities, it doesn't follow that imagination is *inherently* incapable of justifying such claims. And it may turn out that a more sophisticated imagination-based account can be developed that lacks the shortcomings in Yablo's account. As a matter of fact, David Chalmers[54] has developed an imagination-based modal epistemology that appears to be a real advance over Yablo's. In fact, it

---

[54] See, e.g., Chalmers' *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press, 1997), especially ch. 2; and "Does Conceivability Entail Possibility?", in Gendler, Tamar Szabo and John Hawthorne (eds.), *Conceivability and Possibility* (Oxford: Oxford University Press, 2002), pp. 145-200.

appears to solve the dreaded problem of *a posteriori* necessities raised by Kripke and Putnam, which has caused many philosophers to lose a good deal of faith in armchair modalizing.

In this chapter, therefore, I will critically evaluate Chalmers' modal epistemology. First, I'll discuss the problem of *a posteriori* necessities raised by Kripke and Putnam. This will serve the dual roles of motivating Chalmers' account and providing the background for understanding its core elements. After that, I'll offer a brief sketch of the basic contours of Chalmers' modal epistemology. Finally, I'll argue that his account fails to make an advance over our observation-based account, as it is subject to the same criticisms as Yablo's. The present chapter will conclude our evaluation of a representative sampling of accounts of modal epistemology. We'll discuss the implications for Armchair Philosophy (given the truth of the Master Claim) in the next chapter.

3.2 Motivating Chalmers' Account: Lessons from Kripke and Putnam

To help us understand Chalmers' account, it will prove beneficial to examine its roots and primary motivation: *a posteriori* necessities, and the problems they pose for modal epistemology.[55]

A posteriori necessities: a challenge to modal epistemology

Saul Kripke[56] and Hilary Putnam[57] pointed out that many necessary truths about the world are only discoverable *a posteriori* -- away from the armchair, so to speak. So, for

---

[55] Chalmers cites this problem as the motivation for his account in, e.g., "Does Conceivability Entail Possibility?", pp. 156 ff.

example, Hesperus is the same heavenly body as Phosphorus – 'Hesperus' and 'Phosphorus' turned out to be co-referential names for the planet Venus --, and so it's *necessary* that Hesperus is Phosphorus (assuming the necessity of identity). However, it required empirical investigation to discover this fact. And what holds here for individuals also holds for natural kinds. Thus, water and H20 are one and the same stuff, in which case it's necessary that water is H20, and yet it required empirical research to become aware of this.

So there are metaphysical necessities that can't be discerned from the armchair. But if so, then this poses a problem for modal epistemology. For without awareness of a necessary connection between two entities A and B, one is able to conceive of A as existing without B, and vice-versa. In other words, one is able to conceive the impossible, in which case the link between conceivability and possibility is called into question. So, for example, I can conceive of Hesperus as existing without Phosphorus, when in fact (given their identity and its necessity) the state of affairs conceived is impossible. Similarly, a person from an earlier era, or a person who's ignorant of the relevant chemical facts, could conceive of water as existing at a world in which H20 does not, when in fact this is impossible.

---

[56] Kripke, Saul. *Naming and Necessity* (Cambridge, Mass: Harvard University Press, 1980).

[57] "The Meaning of 'Meaning'", in *Mind, Language and Reality: Philosophical Papers, Vol. 2* (Cambridge: Cambridge University Press, 1975), pp. 215-271.

The Kripke/Putnam diagnosis: epistemic counterparts and the Conflation Problem

Thus, Kripke and Putnam taught us that modal error can occur when modalizing in contexts involving *a posteriori* necessities. What's their account of the mechanics of such modal error? Very roughly, it is this: when modalizing in such contexts, one is liable to conflate the target referent or extension with one or more qualitatively identical yet numerically distinct entities.[58] Call any such entity, whether actual or possible – or even merely epistemically possible -- an *epistemic counterpart*. Thus, when I take myself to be imagining water existing without H20, what I'm *really* doing is imagining an epistemically possible world[59] at which there is a natural kind (twater) that's superficially similar to water, but which has a different chemical composition (XYZ). And when I take myself to be imagining Hesperus existing without Phosphorus, what I'm really doing is imagining an epistemically possible world at which there are two *other* heavenly bodies (call them 'Senus' and 'Schmenus') such that one exists without the other. In the first example, twater is water's epistemic counterpart, and in the second example, Senus and Schmenus are Venus' epistemic counterparts.

---

[58] Although Putnam arguably has something like this in mind in his discussion of the Twin Earth case in "The Meaning of 'Meaning'", Kripke explicitly states this sort of account of the mechanics of modal error (at least for *a posteriori* necessity cases) in *Naming and Necessity*. See esp. Lecture III, pp. 141ff. Strictly speaking, Kripke speaks in terms of qualitatively identical epistemic *situations*, which doesn't seem to strictly require qualitatively identical *entities*. However, the difference in terminology doesn't affect matters of substance in our discussion in this chapter.

[59] By "epistemically possible world", I mean a world that the agent in question can't rule out as metaphysically impossible. Thus, on this account, epistemic possibility may outstrip metaphysical possibility.

In short, according to the Kripke/Putnam diagnosis, modal error occurs in *a posteriori* contexts when one conflates the target referent or extension with one or more of its epistemic counterparts. Call this problem for modalizing, *The Conflation Problem.*

A closer look at the Kripke/Putnam diagnosis

What makes The Conflation Problem possible? Putnam is helpful here. According to Putnam, we associate things with what he calls *stereotypes* – salient perceived and behavioral features of a thing or kind learned from ordinary experience.[60] So, for example, our partial stereotype for water is the set of properties denoted by the following description: "the clear, drinkable stuff that falls from clouds, and flows in rivers and streams." Now strictly speaking, Putnam had natural kinds in mind when he introduced his notion of stereotypes. But we can use the notion a bit more loosely for our purposes so as to apply to individuals as well. Given this, we can say that our partial stereotype for Hesperus is the set of properties associated with the following description: "the heavenly body that appears in yonder region of the evening sky". Similarly, our partial stereotype for Phosphorus is given by the following description: "the heavenly body that appears in yonder region of the morning sky".

Now unfortunately, the nature of a thing or kind and its stereotype can come apart: it's at least epistemically possible for a thing's or kind's stereotype to be had by a different thing or kind, and for a single thing or kind to have more than one stereotype. So, for example, H20 and XYZ share water's stereotype, and Venus has both the

---

[60] "The Meaning of 'Meaning'", pp. 227-252. See esp. pp. 249-252.

Hesperus and Phosphorous stereotypes (while Senus and Schmenus split the latter two stereotypes between them).

So there are things and there are epistemic counterparts, and the two share their stereotypes in common, so that entities and stereotypes can be mixed and matched, so to speak. But therein lies the problem. For when it comes to armchair modalizing, we do so on the basis of our folk knowledge of the world. But our folk knowledge of the world is typically stereotype-knowledge. Thus, armchair-based objectual imaginings typically only capture the stereotypes of things and kinds. And it is these facts about entities and their epistemic counterparts that make one liable to conflate the former with the latter. So, for example, one is liable to separate in thought two stereotypes had by a single entity, and then conclude that since the *stereotypes* are separable, their *bearers* are separable. This is a natural explanation of what happened in the Hesperus/Phosphorus and water/H20 cases above

Recap and synthesis

To sum up: Kripke and Putnam pointed out to us that some necessities are only knowable *a posteriori*. But if so, then armchair modalizing in contexts involving *a posteriori* necessities makes one susceptible to the Conflation Problem, i.e., to conflate one's intended referent or extension with one of its epistemic counterparts. And the *mechanics* of the Conflation Problem can be plausibly explained in terms of the following four considerations: (i) armchair modalizing is based on folk knowledge; (ii) folk knowledge is of stereotypes; and (iii) things and kinds share their stereotypes with their epistemic counterparts. And if these things are so, then (iv) the content of armchair modalizing is

77

too coarse-grained to distinguish the former from the latter, thereby making one liable to conflate them.

A useful translation: roles and occupants

It will prove useful in a moment, when we discuss some key elements of Chalmers' modal epistemology, to put our discussion of the mechanics of The Conflation Problem in terms of the role/occupant distinction. Thus, there are the *roles* that individuals and kinds play (e.g., the clear, drinkable stuff that falls from clouds, flows in rivers and streams...), and then there are the individuals and kinds that *occupy* those roles (e.g., H2O, XYZ, etc.). Roles correspond to stereotypes, and occupants correspond to the entities that play them, whether they are the intended referents of modalizing or their epistemic counterparts.

Given the role/occupant distinction, we can explain the Kripke/Putnam problem of *a posteriori* necessities by saying that while the content of armchair modalizing often suffices for individuating *roles*, it's often[61] insufficient for individuating one epistemically possible *occupant* of a role from another. For two or more epistemically possible occupants can share the same role, and a single occupant can play two or more roles. But unfortunately, our armchair knowledge of occupants is typically of their roles. And it is because of this that we are prone to conflate one epistemically possible occupant of a role with another.

---

[61] I say, "*often* insufficient", because phenomenal states might be an exception, as it's hard to see a gap between role and occupant for such states.

78

3.3 A Sketch of Chalmers' Modal Epistemology

Setup: a posteriori *necessities: the moral and Chalmers' solution*

What morals, if any, can be learned from the problem of *a posteriori* necessities? There

are no doubt several, but one will prove especially relevant for the purpose of explicating

Chalmers' modal epistemology: Prior to Kripke's and Putnam's discussions of *a*

posteriori necessities, we thought we were up to the task of modalizing at the level of

occupants – i.e., at the level of particular *bearers* of stereotypes/roles, such as Hesperus

and water. But the Conflation Problem reveals that our powers of armchair modalization

typically penetrate no further than the level of *roles* – the features that things have in

common with their epistemic counterparts.

So *a posteriori* necessities pose a serious problem for armchair modalizing about

particular occupants of roles. However, it's not clear that we should thus throw in the

towel when it comes to armchair modalizing. For from the fact that we're unreliable

when modalizing at the level of particular *occupants*, it doesn't obviously follow that

we're unreliable when modalizing at the level of *roles*. In fact, Chalmers argues that

there's an epistemically safe way to modalize on the basis of role-knowledge. But *how* do

we do this safely? Chalmers' proposed solution to The Conflation Problem involves the

technical apparatus of two-dimensional semantics. However, we need not discuss the

details of that apparatus to understand the fundamental idea behind his account.[62] Rather,

it will prove sufficient for our purposes to give a fairly coarse-grained approximation of

---

[62] See Appendix III for some of the details of Chalmers' two-dimensional semantics and
its role in his modal epistemology.

it. I will therefore express Chalmers' core ideas in intuitive terms below, with the help of the role/occupant distinction.

Answering the Conflation Problem and explicating Chalmers' account

It's helpful to think of Chalmers' modal epistemology as starting from a co-opting of two points from our discussion of *a posteriori* necessities. The first is that armchair knowledge is role-knowledge. So, for example, armchair knowledge of water amounts to knowledge of the watery role – water's stereotype. Second, and relatedly, role-knowledge is coarse-grained; specifically, role-knowledge of a given entity picks out all epistemically possible occupants of a given role. So, for example, role knowledge of water picks out not just H20, but every other epistemically possible occupant of that role (e.g., XYZ), without distinguishing any one occupant from another.

So armchair knowledge is role-knowledge, and role-knowledge picks out all epistemically possible occupants of a given role without distinction. But if these two things are so, then there's a modal epistemic silver lining to the cloud of *a posteriori* necessities. For although there's a sense in which we don't have knowledge of the world before we leave the armchair – we don't know which *occupants* exist at our world without empirical investigation – there is yet an important sense in which we *do* have knowledge of the world from the armchair – we know which *roles* exist in our world. Thus, we don't know from the armchair whether this is an H20-world or an XYZ-world. But we *do* know from the armchair that *something* plays the watery role in the actual world. And as we'll see, there's a way to utilize role-knowledge for armchair modalizing.

80

So armchair knowledge is role-knowledge, and role-knowledge doesn't discriminate between one epistemically possible occupant of a role from another. Now as we've seen, this makes one susceptible to The Conflation Problem. Despite this, however, Chalmers argues that we can modalize from the armchair on the basis of role-knowledge, and yet avoid the Conflation Problem, by distinguishing between two ways of considering a world: as *actual* and as *counterfactual*.[63] When you consider a world <u>as actual</u>, you consider it as *the way the world could actually turn out to be, for all you know a priori*. That is, you single out a set of epistemically possible occupants of a given role, and then consider the hypothesis that they are the occupants of that role at the *actual* world. So, for example, when I consider an XYZ-world as actual, I imagine that the scientists turn out to be wrong, and XYZ is the occupant of the watery role in the actual world.

By contrast, when you consider a world <u>as counterfactual</u>, you imagine it as *a way the world might have been*. That is, you hold a set of occupants of a role fixed – typically, and most saliently, the occupants of that role at the *actual* world – and then consider the world imagined as counterfactual to that world. So, for example, when I consider an XYZ-world as counterfactual, I hold the actual, H20-world fixed, and then consider the XYZ-world as one in which *water* – i.e., *H20* – is XYZ.

So there are two different ways to consider a possible world: as actual and as counterfactual. And the two different ways of considering a world map on to different sets of metaphysically possible worlds that are relevant. When one considers a world as

---

[63] The following discussion is based on Chalmers, *The Conscious Mind*, pp. 56-69, and Chalmers, "Does Conceivability Entail Possibility?", pp. 156-171.

actual, one set of metaphysically possible worlds are relevant – call them

'counteractual'[64] possibilities. And when one considers a world as counterfactual, another

set of metaphysical possibilities – viz., counterfactual possibilities -- are relevant.

Chalmers' modal epistemology: a first pass

Given these points, the basic picture of Chalmers' modal epistemology falls out. His

account has two basic species of conceivability. Call considering a world as actual,

counteractual conceivability, and call considering a world as counterfactual,

counterfactual conceivability. Then the basic account is that counteractual conceivability

is a reliable guide to counteractual possibility, and counterfactual conceivability is a

reliable guide to counterfactual possibility.[65]

Chalmers' account thus distinguishes between two sorts of conceivability-

possibility connections. But given this, it's a short step to solving the Conflation Problem.

For the Conflation Problem occurs when one uses counteractual conceivability as a guide

to counterfactual possibility: for example, when one uses the conceivability of XYZ

watery stuff as a guide to whether XYZ *water* is possible. But as long as one restricts the

use of the counteractual conceivability of XYZ watery stuff to evaluating the

counteractual possibility of XYZ watery stuff, and the counterfactual conceivability of

---

[64] I'm borrowing this term from Stephen Yablo, "Coulda, Woulda, Shoulda", in Gendler
and Hawthorne, eds., *Conceivability and Possibility*, pp. 441-491.

[65] Cf. Chalmers, *The Conscious Mind*, pp. 66-67; Chalmers, "Does Conceivability Entail
Possibility?", p. 165.

82

XYZ water to evaluating the counterfactual (im)possibility of XYZ water, no such conflation can occur.[66]

So far, we've sketched the basic contours of Chalmers' account of when conceivability is evidence for possibility. But it's important to take a closer look at what's involved in these conceivings. For as we've seen in the previous chapters, one's account of conceivability can make or break one's modal epistemology.

Here Chalmers takes his cue from Yablo. For Chalmers adopts Yablo's basic apparatus of objectual imaginings and his two appearance-tests as his method of evaluating modal claims.[67] Thus, on Chalmers' account, one evaluates modal claims via objectually imagining worlds. And a modal claim, "Possibly, P" is *prima facie* justified for one if one is able to objectually imagine a world W such that (i) W appears possible to one, and (ii) P appears to one as true in W.

Thus, Chalmers' modal epistemology is similar to Yablo's, in that both employ objectual imaginings and Yablo's two-clause criterion of modal verification. The main difference is that Chalmers distinguishes between counteractual and counterfactual conceivability, and restricts the proper use of counteractual conceivability to evaluating counteractual possibility-claims. And Chalmers' view appears to be that this innovation is

---

[66] *Ibid.* (both sources and page references).

[67] Chalmers, *The Conscious Mind*, p. 67; Chalmers, "Does Conceivability Entail Possibility?", p. 150 ff. Chalmers mentions his indebtedness to Yablo's account here on p. 150.

sufficient to make things virtually risk-free for armchair modalizing.[68] For such claims can be evaluated solely in terms of the role-content of armchair objectual imaginings. And if so, then not only are they immune to The Conflation Problem, but it appears that such imaginings generate both sorts of appearances for role claims required on Yablo's account: the appearance that the relevant imagined worlds are metaphysically possible, and the appearance that the de-modalized claims are true in those worlds.

3.4 Evaluating Chalmers' Account

What to make of Chalmers' account? It's hard to deny that Chalmers' account makes an advance over Yablo's. For it seems to get around at least one aspect of the Kripke/Putnam problem of *a posteriori* necessities, viz., the Conflation Problem. However, while handling that problem may be necessary for any adequate armchair modal epistemology, there are serious worries that it's not sufficient. In what follows, I will argue that despite the advance gained by Chalmers' account in virtue of its handling The Conflation Problem, it nonetheless suffers from the same sorts of problems that afflict Yablo's account: (i) it can't justify High claims, in which case it makes no advance over our observation-based account in its consequences, and (ii) because of this, his core thesis of imagination as independent evidence of possibility is unmotivated.

---

[68] In at least two places (The *Conscious Mind*, p. 367, n. 32; "Does Conceivability Entail Possibility?", pp. 161-62), Chalmers asserts that Yablo is unwarrantedly tentative about objectual imagining, and that his tentativeness is due to his worries about our proneness to misconstrue the content of our imaginings, as can occur in the *a posteriori* necessity cases that give rise to the Conflation Problem.

84

Chalmers' account can't justify High claims

Let's start with the first problem: Chalmers' account can't justify High possibility-claims. To see this, consider the following claim: Possibly, there is XYZ watery stuff. Now if we are to evaluate this claim via Chalmers' account, it is counteractual conceivability that is relevant. Thus, to evaluate this claim, we are to consider an XYZ world – i.e., a world at which XYZ is the occupant of the watery role -- as actual. And as we've seen, this act does not count as justifying the claim in question unless it involves an objectual imagining of a certain sort. Can we objectually imagine a scenario at which the claim is true?

Well, I can objectually imagine a world at which there is some gently undulating clear liquid in a pool. Furthermore, this world appears metaphysically possible to me; so far, so good. However, does it also appear to me that the claim, 'there is *XYZ* watery stuff', is *true* in that objectually imagined world? No, it doesn't. For again, *all I have to go on* is the content captured by the relevant objectual imagining – the phenomenal imagery of the undulating clear liquid in a pool. True enough: this imagining does make the claim 'there is watery stuff' appear to me as true in that world. But what it *doesn't* get me is an appearance that 'there is *XYZ* watery stuff' is true in that world: for all I'm given in the objectual content of my imagining, the watery stuff imagined is *H20*, and not XYZ. Thus, it fails the second condition of Chalmers' Yablo-style modal verification test.

On the other hand, suppose I just *stipulate* that the watery stuff objectually imagined is XYZ, and not H20. The problem now is that while my imagining passes the

85

second condition of verifying a possibility claim, it fails the first condition. For the XYZ-world now no longer appears to me as *possible*, but as *undecided.*

Objection: modal skepticism and the Popular Argument

Now, as in our discussion of Yablo, one could reply by saying that my doubts stem from an unprincipled and radical form of modal skepticism. Thus, one could generate a version of the Popular Argument against my criticism: if objectual imaginings aren't sufficient to confer *prima facie* justification on the twin water modal claim, then I'm not justified in believing that, say, one can have a justified true belief that isn't knowledge, or even the more humdrum claim that my daughter's ball can get stuck on my roof. But I am justified in believing the latter claims; therefore, objectual imaginings are sufficient to confer *prima facie* justification on the twin water modal claim.

But just as appeal to the Popular Argument cut no ice as a reply on behalf of Yablo's account, so it cuts no ice as a reply on behalf of Chalmers' account. For as we've seen, we have an observation-based account of our knowledge of Low and Middling possibility claims, including not only the humdrum case about my daughter's ball, but also more philosophically interesting modal claims, such as those involved in Gettier cases. However, this account leaves High possibility claims, such as the claim about twin water, unjustified. Therefore, our modal skepticism isn't unprincipled and radical; it's principled and sensibly mitigated. And since this is so, appeal to the Popular Argument fails as a way to resist our criticism of Chalmers' account.

But the problems don't end with twin water. For it seems that the problems spread to all High claims: we have reason to think that High claims *in general* cannot be justified

86

in terms of Chalmers' account. For all such claims go beyond what can be justified in terms of our observation-based account. Therefore, they must be justified in terms of objectual imagining, if his account is correct. But then the dilemma that applied to Yablo's account arises all over again for Chalmers' account: either we allow stipulated content to count as part of the justification-conferring imagining or we don't. If we do, then the stipulations will outstrip the content of what one strictly objectually imagines in a given world, thus undermining the prospects of an appearance of possibility. And if we don't, then what we imagine will be compatible with the falsity of the de-modalized target claim in that world, thus undermining the appearance of the claim's truth in the world objectually imagined. Therefore, it seems that even if Chalmers' account of possibility-knowledge is correct, only Low and Middling possibility claims can be known or *prima facie* justified on his account, in which case it's equivalent to our observation-based account in its consequences.

Chalmers' account is unmotivated

Finally, I want to suggest that as with Yablo's account, we have reason to worry that his core idea – that imaginings provide *independent* evidence for possibility – is undercut. For not only can we account for Low and Middling claims in terms of my observation-based account, but we've just seen that there is good reason to doubt that Chalmers' account can be used to justify High possibility-claims. But if these two things are so, then Chalmers' account is equivalent to my account in terms of the possibility-claims it's able to justify. And if that's right, then like Yablo's account, Chalmers' account fails to provide the best explanation for the epistemic force of the paradigm-case thought

87

experiments. Therefore, pending a better case, we have no reason to think that imagination is an additional source of *prima facie* justification for possibility-beliefs.

The conclusion of the present chapter is that Chalmers' modal epistemology, like Yablo's, takes us no farther than our observation-based account. My claim at the beginning of chapter 1 thus appears to be vindicated: nobody knows whether twin water is metaphysically possible. In the next chapter, we'll bring these results to bear on the broader aims of our inquiry.

Chapter 4:
Modal Skepticism and Armchair Philosophy's Crisis

4.1 Remoteness Modal Skepticism Established as the Default Position

We've spent a considerable amount of time exploring the nature, scope, and limits of our knowledge of metaphysical possibility. It is now time to take stock of that exploration and bring our findings to bear on the broader aims of this dissertation.

In Chapter 1, we looked at some general considerations in favor of Remoteness Modal Skepticism. First, I argued that the Argument from Little Leeway casts doubt on our judgments about High possibilities. Second, I introduced a dilemma for accounts of modal epistemology that attempt to justify High possibility-claims: either they're so permissive as to allow too many false positives, or they're so restrictive that they preclude High possibility-claims from being justified. Third, I sketched an observation-based account of our knowledge of possibility that explains our knowledge of Low and Middling possibilities, while failing to ground knowledge of High possibilities (assuming there are any).

In chapters 2 and 3, we continued our case for Remoteness Modal Skepticism by considering a representative sampling of contemporary accounts of modal epistemology, viz. the accounts of Stephen Yablo and David Chalmers. In those chapters, I argued for two main points. First, neither account vindicates Yablo's abductive argument for imagination as an independent source of *prima facie* justification for possibility-beliefs.

And second, even if one were to grant, at least *arguendo*, that imagination *is* such a source, the problems we raised with Yablo's and Chalmers' accounts indicate that imagination can't take us beyond our observation-based sources and justify claims about *High* possibilities.

Finally, throughout the first three chapters, we saw that appeal to The Popular Argument as a means of resisting Remoteness Modal Skepticism is ineffective. For our positive, observation-based account of possibility-knowledge provides a principled way of accounting for our knowledge of Low and Middling possibilities, while leaving High possibility-claims unjustified.

But if the points above are on track, then it appears that the claim with which we began has been vindicated: no one knows whether twin water is metaphysically possible; *a fortiori* do lack knowledge regarding High possibility-claims even more removed from our knowledge of the actual world. And if that's right, then it appears that Remoteness Modal Skepticism has been established as the default position.

4.2 The Resultant Crisis For Armchair Philosophy

The result of our inquiry so far is a bit of a philosophical buzz-kill. But it gets worse. For if the points above are on track, then it looks as though we have the materials for a formidable challenge to Armchair Philosophy. To see this, recall the argument we met in the Introduction, viz., BAAAP:

> **The Basic Argument Against Armchair Philosophy (BAAAP)**
> 1. Armchair philosophy is legitimate only if thought experiments give us insight into the nature of things.
> 2. Thought experiments give us insight into the nature of things only if

they justify modal claims.
3. Thought experiments don't justify modal claims.
4. So, thought experiments don't give us insight into the nature of things.
5. So, armchair philosophy is not legitimate.

Recall also from the Introduction that the standard non-concessive response to BAAAP is the Modal Epistemology Response, which attacks (3). The basic mode of attack here is to develop and defend an account of modal epistemology that entails that we're capable of having justified High possibility-beliefs. However, if the points of the previous chapters are on track, then to date, this response is a failure; furthermore, the prospects for a successful future response along these lines look bleak.

On the other hand, we've seen that not all is dross in the Modal Epistemology Response. For although there may be good reason to think that knowledge of *High* possibility-claims is out of reach, we've also sketched an observation-based account that entails solid epistemic grounding for *Low* and *Middling* possibilities. In fact, we've seen that Low and Middling possibility knowledge is all that's required for a good many thought experiments, including the Gettier cases. But if so, then we have reason to think that a number of philosophically interesting claims can be evaluated by means of such knowledge. Thus, our observation-based account provides a solid basis for a qualified rejection of (3).

However, this response only *mitigates* the force of BAAAP; it does not *eliminate* it. For while our observation-based account may well be able to ground our knowledge of many philosophically interesting claims, it's arguable that a good many such claims cannot be so grounded. Think, for example, of the large host of thought experiments that depend on High claims: claims about, e.g., the possibility of zombies and disembodied

91

souls in the philosophy of mind literature; personal fission in the personal identity literature; Thomson's people seeds in the abortion literature; Davidson's swampman in the philosophy of mind and language literature; etc. Thus, our account offers little encouragement that such thought experiments support their corresponding possibility claims. And if that's right, then assuming the truth of The Master Claim, we'll have to look beyond The Modal Epistemology Response if a non-concessive response to BAAAP is to be carried any further.

4.3 Exploring An Untried Response: Non-Modal Thought Experiments

A moral of the first three chapters is that there is pressure to think that one can do Armchair Philosophy, and one can investigate the nature of things in the world, but one often cannot do the latter by means of the former. However, I will argue in the next two chapters that there is a way between the horns of this dilemma. To see this, recall that BAAAP followed from three basic premises. Now we've seen that the standard responses were: The Naturalized Philosophy Response, which accepts all the premises; the Conceptual Analysis Response, which rejects (1); and the Modal Epistemology Response, which rejects (3). Thus, interestingly, all the camps at least *de facto* accept (2). Now recall that we're assuming, for the purposes of this dissertation, the truth of The Master Claim, viz., that thought experiments tell us about the world. We will therefore leave to the side a consideration The Naturalized Philosophy Response and The Conceptual Analysis Response. But if (2) isn't unassailable, then there's room for a different response to BAAAP -- a form of armchair philosophy that at least partly relies

on thought experiments that have the following two features: (i) they tell us about the world, but (ii) they do so without reference to *modal* facts.

In Chapter 5, I will give an account of thought experiments that satisfy the two conditions just mentioned. The key claim of that chapter will thus be that there is at least one type of thought experiment – what I call the *non-modal thought experiment* -- whose epistemic force does not depend upon whether it justifies modal claims. In Chapter 6, I'll offer evidence to suggest that non-modal thought experiments can do the bulk of the work previously assigned to thought experiments that rely on High possibility-claims. This will provide the basis for a qualified rejection of (2) of BAAAP.

In this way, I will complete my strategy for responding to BAAAP. And while, as we've seen, there are other, concessive ways of responding, given the shortcomings in The Modal Epistemology Response discussed above, the one I defend is arguably the only one that allows for a substantive connection between armchair theorizing and the world.

CHAPTER 5:
Non-Modal Thought Experiments

At the end of the previous chapter, we saw that while BAAAP poses a significant threat

to Armchair Philosophy as a means of gaining information about the world, this threat

can be averted if thought experiments can yield such information without crucial appeal

to other possible worlds. We also saw that there is logical space for such an account. The

main goal of the current chapter is to provide a sketch of an account of non-modal

thought experiments, and to begin to argue for its plausibility.

5.1 Non-modal Thought Experiments Characterized

The basic idea

Throughout most of our discussion of modal epistemology in the first two chapters, we

spoke as though all thought experiments were the same in a crucial sense, viz., that they

all are aimed at depicting other possible worlds, and indeed dependent for their epistemic

force upon whether they *succeed* in doing so. Call any such thought experiment a *modal*

thought experiment. In this chapter, however, I will argue against this assumption by

explaining and defending the existence of thought experiments of another sort: non-

modal thought experiments. As I characterize it, a *non-modal thought experiment* is one

whose epistemic force does not depend upon whether it justifies a modal claim at which

the relevant entities at issue come apart. The notion of a non-modal thought experiment

94

will become clearer as we discuss particular cases, but a characteristic feature of non-modal thought experiments is their ability to help us discern *metaphysical* facts without appeal to *modal* facts.

I will attempt to flesh out the notion of a non-modal thought experiment in a moment, but to get an intuitive sense of the basic idea I'm advancing, consider the following two cases:

> Case 1: The Spinoza-world:
> We discover that Spinoza was right: our world is a necessary being, and has all and only the features it has of metaphysical necessity. However, in our world – the only world -- there is no instance of justified true belief without knowledge, and no instance of knowledge that isn't justified true belief. Still, even if this turns out to be so, it seems that the epistemic force of (for example) Gettier thought experiments would be unaffected by such a discovery, in that they would still show that justified true belief isn't knowledge.

> Case 2: The god with a hang-up:
> We discover that there are many possible worlds – indeed, infinitely many.   However, we also discover that an Anselmian god exists, and he has a certain hang-up that's essential to him: he wants every case of justified true belief to also be an instance of knowledge, and every instance of knowledge to be a case of justified true belief.[69] Therefore, he makes sure that knowledge and justified true belief march in lockstep across all possible worlds in which either one exists. Still, even if this turns out to be so, it seems that the epistemic force of Gettier thought experiments would be unaffected by such discoveries, in that they would still show that justified true belief isn't knowledge.

Although there are interesting differences between the two cases here, they share a common recipe: (i) we consider a particular philosophical analysis, or at least partial analysis, of some entity, (ii) we suppose that the properties involved in the *analysans* and

---

[69] In much the same way that the moral perfection of Descartes's omnipotent and omniscient god ensures that he makes clear and distinct perception to be an infallible criterion of truth.

the *analysandum* have the same modal profile, but (iii) despite (ii), we have the intuition that the thought experiment yet reveals the analysis to be inadequate. Thus, these cases at least suggest that thought experiments can justify substantive claims without relying on the well-worn inferential route from conceived separability to possible separability to distinctness between entities. How can this be? Although I'll need to develop the point later, I want to suggest that it's because thought experiments can at least sometimes get at factors relevant to the truth-values of philosophical claims that are more fine-grained than what can be individuated by means of possible worlds. But if so, then while separating two or more entities at a possible world – whether in ours or some other – is sufficient for noticing their distinctness, it isn't always required. And if that's right, then the door is at least ajar for the possibility that thought experiments can aid us in noticing distinctness without appealing to the familiar device of separation at a possible world.

5.2 Arguing for the Legitimacy of Non-Modal Thought Experiments

With the basic idea on the table, we can now begin to argue for it and develop it.

First consideration: reductio *and* per impossibile *thought experiments*

I think the cases mentioned above provide at least some intuitive evidence that thought experiments need not make crucial reference to other possible worlds. Here is another piece of evidence: we're already familiar with two sorts of thought experiment that don't appeal to possible worlds in order to justify various philosophical claims. For, first, there are *reductio* thought experiments, such as the one related to the Grandfather Paradox. With this thought experiment, one imagines going back in time and killing one's

96

grandfather. But if we hold certain assumptions fixed, the scenario is metaphysically impossible. For it includes the incompatible states of having been born and having never been born. The key thing to notice here is that the impossibility in question is no hindrance at all to the force of the thought experiment, for the contradiction is precisely the thing that gives it its force. Now the point is *not* that this particular thought experiment is successful; it may or may not show that time-travel is impossible. Rather, the point is one about the *method of reasoning* involved. In particular, it illustrates how thought experiments can have philosophical utility in a way that doesn't depend on justifying claims about *possible* worlds.

Another kind of useful thought experiment that makes no appeal to possible worlds is the *per impossibile* thought experiment. Like *reductio* thought experiments, these also appeal to a putative metaphysical impossibility. So, for example, the medievals tried to discern the dependence-relationship between God and abstracta by having us consider whether abstracta would exist if, *per impossibile*, God did not exist. Again, as with *reductio* thought experiments, *per impossibile* thought experiments involve a metaphysical impossibility, and yet they have philosophical utility in virtue of their capacity to support or defeat philosophical claims.

Now while these sorts of thought experiments are interesting – indeed, it would be interesting to explore the extent of their utility by seeing how many thought experiments could be re-construed as *reductio* and *per impossibile* thought experiments – I here set them aside to consider some (perhaps) hitherto unrecognized kinds of thought experiments. Before doing so, however, I want to underscore the main point here, which

97

is that there is precedent for thought experiments that have philosophical utility without appeal to *possible* worlds. Next, I will argue that thought experiments can have such utility without crucial appeal to worlds of *any* sort, whether possible or impossible.

Second consideration: the Twin Earth thought experiment and the inconsistent set

Consider the following variation on Putnam's Twin Earth thought experiment:

> **Twin Earth**:
> There is another planet qualitatively identical to ours – call it *Twin Earth* -
> - except that XYZ plays the watery role on Twin Earth. Now if the
> meaning of 'water' = descriptive content about the watery role, then
> 'water' refers to *both* the H20-tokens on Earth *and* the XYZ-tokens on
> Twin Earth. But this isn't right: 'water' refers to just the H20-tokens when
> uttered by Earthians, while it refers to just the XYZ-tokens when uttered
> by Twin Earthians; therefore, the meaning of 'water' ≠ descriptive content
> about the watery role.

Now suppose the case for Remoteness Modal Skepticism in the first three chapters is persuasive. If so, then there is pressure to think that if the force of the thought experiment here depends upon whether it accurately depicts a possible world, it's a hopeless failure. For the claim about the metaphysical possibility of an alien substance that perfectly plays the watery role is a modal claim that cannot be traced back to our knowledge of the actual world: I've never observed such an alien substance; nor have I received legitimate testimony as to its existence; therefore, I cannot generate an argument from types to tokens for its possible existence; nor can our folk theory of the actual world cannot support belief in its possible existence; and such a substance isn't relevantly similar to things in the actual world, in which case I'm unable to justify its possible existence via an argument from analogy. Therefore, the claim about the possible existence of non-H20 watery stuff lacks sufficient justification to merit acceptance.

98

So there is pressure from the case for Remoteness Modal Skepticism to say the thought experiment fails to justify the *modal* claim about the possibility of non-H20 watery stuff. On the other hand, it's hard to resist saying that the thought experiment shows *something*, viz., that the descriptive content about the watery role associated with 'water' isn't sufficient to determine reference. But how can this be if the thought experiment fails to justify a *modal* claim?

To help us get clearer on the problem at hand, consider the following set of propositions:

1. Thought experiments can reveal metaphysical facts.
2. Thought experiments can reveal metaphysical facts only if they reveal modal facts.
3. Thought experiments can't reveal modal facts.[70]

This is an inconsistent set, and so we must reject at least one of its propositions. Let's look at each option in turn.

First, we could reject (1). According to this response, thought experiments aren't legitimate guides to the natures of things. However, this isn't obviously right. For consider again our variation on Putnam's Twin Earth thought experiment. *Prima facie*, the thought experiment shows that the descriptive content about the watery role is not sufficient for determining the referent of 'water'. But if so, then there is pressure to hold on to (1).

Alternatively, one could reject (3). According to this response, thought experiments yield significant modal facts. However, in chapters 1-3, we saw that there is

---

[70] For simplicity's sake, I omit qualifications that reflect our knowledge of Low and Middling modal facts, as High modal facts are at issue here.

pressure to accept Remoteness Modal Skepticism as the default position, and that the prospects for moving beyond that position to something more substantial do not inspire confidence. So there is pressure to accept (3).

Thus, there is pressure to hold on to both (1) and (3). But if so, then there is pressure to reject (2). In other words, there is pressure to say that at least some thought experiments are able to tell us about metaphysical facts *even if they can't tell us about modal facts*.

5.3 Obstacles to Non-Modal Thought Experiments

The obstacles: possible-worlds analyses of strict implication and of essence

I have argued that there is not only logical space, but also a non-trivial basis, for thinking that non-modal thought experiments can yield information about the world. However, there are at least two major obstacles to accepting non-modal thought experiments: (i) possible-worlds analyses of strict implication, and (ii) possible-worlds analyses of essential properties. These are obstacles because a number of philosophical positions are spelled out in terms of strict conditionals and in terms of claims about the essence of a property or thing. But if so, then if the notions of strict implication and essence are spelled out in terms of truth across possible worlds, then it would seem that the only way to defeat claims spelled out in these terms is to appeal to possible worlds at which they are false. And if so, then if a thought experiment is to be useful in the task of evaluating or defeating such claims, then it must help us to justify the existence of such defeater-worlds.

100

In this section, therefore, I will try to go some way toward overcoming these obstacles. My strategy will be to sketch some reasons for thinking that possible-worlds analyses of strict implication and of essential properties are inadequate, as well as to sketch alternative analyses of these notions that lack the problems of their modal counterparts. Our inquiry here will also provide the materials for a positive account of non-modal thought experiments.

Overcoming the first obstacle: Roy's relevant semantics for entailment

The standard account of the semantics for strict implication spells out the relation between antecedent and consequent in terms of possible worlds: P strictly implies Q just in case there is no possible world at which P is true and Q is false. Thus, 'P strictly implies Q' is commonly symbolized as '$\Box(P\rightarrow Q)$', where '$\Box$' denotes the necessity operator of modal logic. So, for example, consider the following strict conditional:

1. Necessarily, if Steve's a bachelor, then Steve's unmarried.

There is no possible world at which Steve is both a bachelor and married; therefore, according to the standard account of strict conditionals, (1) is true, and its antecedent *entails* its consequent.

So far, so good: the standard account accords with our intuition that Steve's being a bachelor entails Steve's being unmarried. Furthermore, its explanation for *why* (1)'s antecedent entails its consequent accords with our intuitions as well: Steve's being a bachelor "necessitates" Steve's being unmarried. However, a number of philosophers have pointed out that the standard possible-worlds account of entailment has counter-intuitive results. So, for example, consider the following strict conditional:

101

2. Necessarily, if all bachelors are unmarried, then $1+1 = 2$.

In (2), both the antecedent and the consequent are true at all possible worlds. If so, then there is no possible world at which the antecedent is true and the consequent is false. So by the standard account of the semantics of strict implication, (2) is true, and so the antecedent entails the consequent. But this case brings out a counterintuitive implication of the standard account. For it's counterintuitive to say that A *entails* B if A and B aren't relevantly related – i.e., if A's obtaining is *irrelevant* to B's obtaining.

The problem of irrelevance can be brought out even more starkly by considering the following proposition:

3. Necessarily, if cats are fish, then $1+1=2$.

Now some might quibble about whether there is a possible world at which cats are fish, but few will argue that there is a possible world at which the sum of 1 and 1 is something other than 2. But if there is no possible world at which (3)'s consequent is false, then there is no possible world at which its antecedent is true and its consequent is false. Therefore, according to the standard account of the semantics of strict conditionals, (3) is true, and so its antecedent entails its consequent. But it's counterintuitive that one thing *A* should count as *entailing* another thing *B* if *B* always obtains *without A*; *a fortiori* is the entailment claim counterintuitive if it's *necessary* that *B* always obtains without *A*! But if the standard account has these counterintuitive implications, there is pressure to think that the standard account doesn't capture our intuitive notion of entailment, however helpful the standard account may be in other respects.

102

We've just seen that there are some good reasons to think that the standard account of strict conditionals isn't terribly sensitive to relations of relevance. But if not, then the worry is that they're not ideally suited to characterize full or partial analyses of entities. But if not, then to the extent that thought experiments aim to evaluate such analyses, standard strict conditionals aren't well-suited for the task.

But if the standard possible-worlds account is inadequate, what account is adequate? Unfortunately, there is as of yet no widely agreed-upon account on offer. However, it would clearly go far beyond the scope of this dissertation to adequately address this issue. In light of the difficulty and magnitude of the task of developing and defending "the one true account of entailment" (if there is such a thing), I'd like to put that task to the side and pursue the more modest and manageable goal of sketching an account of conditionals that is suggestive and not obviously implausible, and at least sheds light on our intuitions of the utility of non-modal thought experiments. Thus, my goal here is not to argue that the following account should *replace* the standard possible-worlds account. Rather, it is the more modest one of pointing to an *additional*, not-implausible way of construing entailment claims when we evaluate them by means of non-modal thought experiments. If I can do this much, then I can get around the obstacle that the standard account poses for non-modal thought experiments. For our purposes, then, it is enough to sketch the rudimentary ideas of an alternate account, and then to explain how non-modal thought experiments can be used to evaluate them.

Tony Roy has offered an intriguing relevant-semantics account of entailment that appears to avoid the aforementioned problems that apply to the standard possible-worlds

103

account.[71] On Roy's account, entailment claims, when relevance is factored in, are explained in terms of *property structures*, which are finer-grained and more abundant than *worlds*.[72] If so, then we have a way of explaining why thought experiments succeed without appeal to other possible worlds: they indicate the relationships among property structures, and these relationships can be revealed in a thought experiment even if it's impossible, or inscrutable whether it's possible, that a possible world can sustain such a structure. Or to put it another way: On Roy's relevance-logic analysis of entailment, the truth-conditions of entailment claims are *independent* of claims about possible worlds.

According to Roy's account, logical facts supervene not upon possible worlds, but upon property structures.[73] Now a theory of properties must meet at least the following two conditions if properties are to play the roles required of them in Roy's account: (i) They are independent elements that can be combined in arbitrary ways, and (ii) there are no restrictions on which properties, and how many, may be combined: properties are capable of being combined even if their combination would render them uninstantiable.[74] Condition (i) allows that properties can be conjoined to form conjunctive properties, disjunctive properties, etc., and of arbitrary "length". And condition (ii) allows the

---

[71] The following discussion is a summary of the relevant points in Tony Roy's "Making Sense of Relevant Semantics" (ms.), esp. pp. 1-13. The paper can be found online at: <http://philosophy.csusb.edu/~troy/msrs-paper.pdf>

[72] *Ibid.*, pp. 7-9.

[73] *Ibid.*, p. 11. Roy allows that structures of other sorts might be capable of playing the roles he ascribes here to properties. For the sake of convenience, though, I will follow Roy and explain his account in terms of properties.

[74] *Ibid.*

104

account to be more fine-grained than any account spelled out in terms of possible worlds, as it allows for uninstantiable property structures – property structures that are instantiated in no possible world; a structure that sinks like a stone on the sea of actuality is no less of a structure for that. As we will see, the fine-grainedness of Roy's account provides a way to understand the legitimacy of non-modal thought experiments.

The most salient part of Roy's theoretical underpinnings for his account of entailment is his notion of a *closed* property. To get this idea before us, it will help if we first consider a set S of properties formed in the following way. Say we have a property, F. Now suppose F isn't an island unto itself, as it were, but instead inherently involves the instantiation of other properties. Thus, suppose the instantiation of F inherently involves the instantiation of A, B, and C. Now let set S be formed by taking just A and the properties intrinsically involved in A's instantiation. Then S looks like this:

$$S = \{F, A, B, C\}$$

Now consider the property W composed of exactly the members of S. Then by Roy's definition, W is a closed property.

Now if it turns out that the instantiation of these other properties (i.e., A, B, and C) *themselves* involve the instantiation of further properties, then *these* are included in S as well. Thus, suppose the instantiation of A inherently involves the instantiation of $a1$, $a2$, and $a3$; the instantiation of B thereby involves the instantiation of $b1$, $b2$, and $b3$; and the instantiation of C thereby involves the instantiation of $c1$, $c2$, and $c3$. Then S would instead look like this:

$$S = \{F, A, B, C, a1, a2, a3, b1, b2, b3, c1, c2, c3\}$$

And closed W would be the complex property composed of its members. And so on for however many other constituents are intrinsically involved in the instantiation of F.

We are now in a position to understand the notion of a closed property: a property W is *closed* iff W includes some property F and all the properties $f1, f2, f2, ..., fn$ that are thereby instantiated in virtue of F's instantiation.[75]

With the addition of Roy's notion of a closed property to his theoretical picture of arbitrarily combinable properties, it looks as though we have a suitable supervenience base for entailment, which, for convenience, I will call *Roy-Entailment* (RE):

(RE) P entails Q iff no closed *being W* has *being P* as a constituent

without having *being Q* as a constituent.[76]

On this account, then, entailment is a relation that supervenes on closed property structures. And the reason why entailment from P to Q holds across the class of closed property structures is that *being Q* is a <u>sub-complex</u> of *being P*.[77] So, for example, it's intuitive that 'x is round' entails 'x is shaped'. On Roy's account, the entailment relation here supervenes on something much "thinner," and much more abundant, than possible worlds, viz., the set of closed properties. That is, there is no closed property that has *being round* as a constituent that does not also have *being shaped* as a constituent. And

---

[75] Roy states his definition of a closed property more formally as follows: "Say a set w of such properties is closed when any $v \subseteq w$ is such that if a thing instantiates the properties in v, then w includes also any properties thereby instantiated by the thing." *Ibid.*

[76] *Ibid.*, p. 11. Roy labels this notion of entailment, *LV*.

[77] As noted in *ibid.*, the term, 'sub-complex', and its usage here, is based on Michael Jubien's in his "Actualism and Iterated Modalities", *Philosophical Studies* 84 (1996), pp. 109-125.

this is explained by the fact that *being shaped* is a <u>sub-complex</u> of – i.e., is an inherent part of the internal structure of -- *being round*.

To be sure, Roy's strict conditionals are inherently related to the possible-worlds analysis of necessity, in that any true Roy-style strict conditional is true at all possible worlds. However, Roy's account doesn't require that the converse holds: it allows for the epistemic possibility that Q is true whenever P is true in all possible worlds, and yet P does not strictly entail Q in Roy's sense. Rather, on Roy's account, facts about sub-complex relations are more fundamental than the modal facts here, in that the former *explain* the latter, as the latter *supervene* on the former: Q is true whenever P is true in all possible worlds *because* Q is a sub-complex of P.

In short, on Roy's relevant semantics, entailment supervenes on closed property structures. But if so, then we have an account of entailment that's much more fine-grained than a possible-worlds account. And if that's right, then we have a way of making sense of non-modal thought experiments. For if we can make sense of whether P is sufficient for Q and of whether P is necessary for Q by appeal to entities more fine-grained than worlds (viz., closed properties), then the door is open for the possibility of an account of thought experiments that makes no crucial appeal to possible worlds. And the proposal here is that *thought experiments can reveal facts about sub-complex relations (or the lack thereof) between constituents of closed properties.*

Roy's account also provides a way to explain our intuitions in the Spinoza-world case and the God-with-a-hang-up case: even if there is only one possible world -- or even if there are many worlds, but god holds fixed certain arbitrary property configurations

across worlds -- there is yet a realm of properties that transcends the worlds, and serve as their fundamental constituents. And while the *modal* facts can be affected by factors such as those just mentioned, the realm of properties are such that certain members bear internal, sub-complex/whole relations to certain others, and so these cannot be altered -- they are unaffected by what goes on in the worlds. And if that's right, then it's possible, at least in principle, to know facts about property structures in a way that's independent of knowledge of modal facts.

In light of this partial sketch of Roy's relevant semantics, consider again Putnam's Twin Earth thought experiment: we imagine a scenario according to which H20 plays the watery role on Earth, but XYZ plays the watery role on Twin Earth. Given this, it would seem that 'water' refers to all and only H20-tokens when uttered by Earthians, and it refers to all and only XYZ-tokens when uttered by Twin Earthians. But if the description associated with 'water' determined reference, we should conclude that 'water' refers to *both* the H20-tokens *and* the XYZ-tokens. Thus, a term's associated description doesn't determine its reference. Now suppose it turns out that there is no metaphysically possible world at which some non-H20 stuff plays the watery role. Should we then conclude that Putnam's thought experiment fails to refute descriptivist theories of reference? No. Rather, it seems that we should think that it succeeds *irrespective* of whether such a world is possible. And now that we have Roy's account of entailment in hand, we have a way to explain this: the Twin Earth thought experiment enables us to see that, instantiable or not, there is a closed conjunctive property structure, *being watery* and *being XYZ.* But if so, then there is at least one closed property structure

108

that has the properties associated with the descriptive content of 'water' that lacks the property, *being water*. And if that's right, then the latter properties don't "entail" the former properties in the way that Roy's version of relevant semantics defines property entailment.

Overcoming the second obstacle: Fine's account of essence

We've seen that the first main obstacle to non-modal thought experiments – viz., possible-worlds analyses of strict implication – isn't so formidable. However, as I mentioned earlier, another obstacle remains: possible-worlds analyses of essential properties. I therefore need to advance a non-modal analysis of essence. But I suspect that some will have a problem with this account. For it relies on the idea that a property can be necessarily coextensive with a thing's essential properties without *itself* being essential to the thing. But how can this be? Isn't it part of philosophical orthodoxy that a property is essential to a bearer just in case the bearer has the property in all possible worlds in which it exists? While I agree that it is indeed orthodoxy, I think that modal accounts are nonetheless inadequate. I'll briefly discuss two reasons for thinking so: (1) the modal account admits of counterexamples, and more fundamentally, (2) the modal account is too coarse-grained to adequately differentiate essential from non-essential properties. Let's discuss these criticisms in turn.

Alvin Plantinga's account of essential propertyhood is representative. In *The Nature of Necessity*, Plantinga tells us that a property, *P*, is essential to its bearer, *x*, if and

109

only if x is P in all possible worlds in which x exists.[78] So, suppose that Ralph has the property of being a Californian in some possible worlds, but not in others. On Plantinga's analysis, then, being a Californian isn't essential to Ralph. By contrast, suppose that Ralph has the property of being a person in all possible worlds in which he exists. In this case, Plantinga's account would entail that being a person is essential to Ralph.

I think Kit Fine provides a nice counterexample to Plantinga's analysis in his 1994 paper, "Essence and Modality".[79] Thus, consider the set whose only member is Socrates, and call that set 'singleton Socrates'. There is no possible world in which Socrates exists yet lacks the property of being the member of singleton Socrates. So, by Plantinga's modal definition of essence, being a member of this singleton set is essential to Socrates. But intuitively, being a member of this set is *not* essential to or constitutive of Socrates. As Fine puts it, "Strange as the literature on personal identity may be, it has never been suggested that in order to understand the nature of a person one must know to which set he belongs. There is nothing in the nature of a person...which demands that he belongs to this or that set or which even demands that there be any sets."[80]

Another sort of counterexample can be generated by appeal to world-indexed properties.[81] So, for example, consider the brownness of my coffee table. *Prima facie*,

---

[78] (Oxford: Oxford University Press, 1974). This is a slight paraphrase of Plantinga's account on p. 54.

[79] *Philosophical Perspectives*, Vol. 8, Logic and Language (1994), pp. 1-16.

[80] *Ibid*, p. 5

[81] This counterexample was suggested to me by Peter J. Graham.

that's a contingent, accidental property of my coffee table. But now consider the world-indexed property of *being brown-in-alpha*, where 'alpha' denotes the actual world. Then my coffee table has the property of being brown-in-alpha at all the worlds in which it exists. But if so, then according to the modal analysis, being brown-in-alpha is an *essential* property of my coffee table. But *prima facie*, being brown-in-alpha isn't part of *what it is to be* my coffee table, any more than being the only member of singleton Socrates is part of what it is to be Socrates. Thus, the standard modal analysis of essential properties appears to be inadequate, as it admits of counterexamples such as these.[82]

This brings me to my second complaint against Plantinga's modal account of essential properties: it's too coarse-grained. For in Fine's case of Socrates and his singleton set, the set of worlds containing Socrates and the set of worlds containing singleton Socrates are the same set. Therefore, on the modal analysis of essential properties, one can't make sense of the intuitive asymmetry here: it is part of the nature of singleton Socrates that it contains Socrates as a member, but it's no part of the nature of *Socrates* that he be a member of singleton Socrates. Something has gone wrong here, and it is the adoption of a modal analysis of essence. As Fine's case nicely illustrates, possible worlds are just too coarse-grained to do the work of singling out essential properties.

But if modal analyses are inadequate, what sort of analysis *is* adequate? Offering a full-blown positive account of essence is of course beyond the scope of this dissertation, but I would like to offer, in at least broad outline, a sketch of an alternate,

---

[82] I offer another counterexample to modal analyses of essential properties in chapter 6.

non-modal account. Here again, Fine's article is helpful. Fine distinguishes between *modal* and *definitional* accounts of essential properties. Modal accounts are of the sort Plantinga gives. Definitional accounts are of the sort that Aristotle gave. Thus, just as one can give the definition of what a word *means*, so one can give the definition of what an object – or kind of object – *is*. As Aristotle put it, "clearly, then, the definition is the formula of the essence." (*Metaphysics* 1031a12).[83]

To put the ideas in a slightly different way: on both modal and definitional accounts of essence, essential properties are inextricably tied to modal properties. Thus, just like the modal analysis of essence, definitional accounts entail that if F is essential to x, then x is F in all possible worlds in which x exists. But the difference is that on the definitional account, the relationship between modal properties and essential properties is *asymmetric*: the modal properties supervene on the essential properties, but the essential properties do not supervene on the modal properties. Thus, the converse of the conditional above holds for the modal account of essence, but not for the definitional account – x might be F in all possible worlds in which x exists, but for all that, F might not be essential to x (hence the cases of Socrates and singleton Socrates, the coffee table and its alpha-indexed brownness, etc.).[84]

---

[83] Both the reference and the definitional construal of Aristotle's account of essence are Fine's. See *ibid*, p. 12.

[84] I defer discussion of the semantics for essence claims of this sort to chapter 6.

112

5.4 Roy-Conditionals, Fine-Analyses, and Non-Modal Thought Experiments

It's important to pause for a moment here and emphasize an important implication of the truth-conditions for Roy-conditionals and for definitional accounts of essence: their truth-conditions *do not make essential reference to other possible worlds*. There is a good reason for this: *a non-constituent property can be necessarily co-extensive with a constituent property*. So, for example, the property of having three sides is necessarily co-extensive with the property of having three angles (modulo certain qualifications about, for example, being a *closed* figure), but the former isn't a constituent – doesn't enter into the being – of the latter. Or consider Descartes' account of epistemic justification in terms of clear and distinct perceptions of the intellect. On this account, it isn't *essential* to clear and distinct perceptions that they are true. For according to Descartes, if it weren't for an all-powerful god to ensure the connection between clear and distinct perceptions and true beliefs, the former wouldn't be an infallible indicator of the latter. However, Descartes' god ensures that the two are necessarily co-extensive. So again, being an essential property is one thing, being a necessarily co-extensive property is another, and the latter isn't an infallible indicator of the former. But if that's right, then the methods of evaluation for Roy-conditionals and Fine-analyses may diverge from those of strict conditionals. The moral here is that constituent claims and analyses can be evaluated and defeated without appealing to other possible worlds. And if that's right, then the door is open for evaluating them *via non-modal thought experiments*.

113

5.5 Recap

Let's recap the territory we've covered in this chapter. I've been arguing that there is room for a hitherto-overlooked form of philosophical thought experiment, viz., *the non-modal thought experiment*. I've argued that *reductio* and *per impossibile* thought experiments appeal to impossibilities to support philosophical claims, in which case justifying possibility claims isn't a necessary feature of thought experiments. I also argued that the case for Remoteness Modal Skepticism, when combined with the resilient intuitive force of certain thought experiments, generates pressure to think that the force of at least some thought experiments don't depend for their force upon whether they justify a modal claim. However, we saw that at least two main obstacles stood in the way of accepting the legitimacy of non-modal thought experiments: possible-worlds accounts of strict implication and possible-worlds accounts of essential properties. But we've seen that there are reasons to think that these obstacles aren't so formidable, as they have implausible implications, and that alternative, finer-grained accounts of these notions are available. In fact, these latter accounts provide an explanation for *why* non-modal thought experiments are legitimate. Therefore, if these points are on track, then it looks as though non-modal thought experiments have some sturdy legs to stand on.

CHAPTER 6:
The Framework

It is now time to synthesize the results of the previous chapters and construct an account

of the variety and utility of thought experiments. This will primarily involve defining and

explaining a set of elementary concepts and distinctions. In particular, I'll do three things

here: (i) Distinguish between types of thought experiments; (ii) distinguish between types

of philosophical claims; and (iii) point out the types of thought experiments appropriate

for each type of philosophical claim. This will constitute a framework for doing Armchair

Philosophy – or at least a significant fragment of it – in a way that avoids the problems

raised by Remoteness Modal Skepticism. With the framework thus laid out, I'll show

how it can be applied to make progress in the doing of Armchair Philosophy, and how the

framework itself can solve some basic issues in metaphilosophy.

6.1 Kinds of Thought Experiment

High-flying vs. low-flying thought experiments

Recall that in chapters 1-3, we argued that our knowledge of possibility is limited to what

can be properly inferred from our knowledge of the actual world. We also saw a variety

of ways in which our knowledge of possibility traces back to our knowledge of actuality:

(i) to deductions of possibility from observations of actuality; (ii) to various sorts of

inferences from tokens to types of observed actualities; (iii) to our folk theories of how

115

the actual world works; (iv) to arguments from relevant similarities with the actual world; and (v) to testimony of all the above.

In light this account, we can distinguish between two sorts of thought experiment: *high-flying* and *low-flying*. Thus, let's say that low-flying thought experiments are those that can be grounded in our knowledge of the actual world in ways (i)-(v) listed above. Thus, I can imagine scenarios according to which my coffee table exists a foot to the left of its current position, my car is painted pink, and my computer crashes. Furthermore, I know these scenarios are metaphysically possible, since I know via (e.g.) my folk theory that ours is a world in which tables can be moved about, cars can be painted different colors, and computers can crash. These imagined scenarios are therefore low-flying thought experiments. Such cases are fairly trivial, but as we saw in chapter 1, many are more interesting. For example, certain Gettier cases are low-flying, as well as, for example, Locke's locked room case, certain trolley problem cases, and (arguably) Thomson's violinist case.

In contrast to low-flying thought experiments are high-flying thought experiments. High-flying thought experiments are those that cannot be grounded in our knowledge of the actual world in ways (i)-(v) listed above. Thus, I can generate thought experiments according to which I exist apart from my body, an Anselmian god exists in some possible world, and a wizard works his protective magic whenever his favorite glass object is about to shatter. However, I've never observed such things; nor have I received legitimate testimony as to their existence. Therefore, I cannot generate an argument from types to tokens for their possible existence. Furthermore, our folk theory

116

of the actual world cannot support their possible existence. Finally, I have no reason to think such things are relevantly to things in the actual world, and thus I'm unable to justify their possibility via an argument from analogy. Therefore, such thought experiments are of the high-flying variety.

Modal vs. non-modal thought experiments

So far, we have divided thought experiments into two sorts: high-flying and low-flying. I would now like to divide them further, in terms of whether they are *modal* or *non-modal*. This distinction is based on our discussion of non-modal thought experiments in the previous chapter. Thus, let's continue with our use of that notion and say that modal thought experiments are those whose utility depends on whether they justify a modal claim – most saliently for our purposes, a claim that something is metaphysically possible. The Cartesian thought experiment involving conceiving of yourself existing apart from your body is a paradigm case of a modal thought experiment. It's supposed to justify the modal claim that it's metaphysically possible for your mind to exist apart from your body. And if it doesn't justify that such a thing is possible, it is thereby deemed a failure – if it can't justify a *modal* fact, then it can't justify a *metaphysical* fact.

By contrast, non-modal thought experiments are those whose utility does not depend on whether they justify modal claims. Thus, unlike modal thought experiments, they can justify the existence of metaphysical facts without having to first justify some modal fact. Such thought experiments have a variety of functions, but they typically involve showing that a given philosophical analysis of a property is inadequate. Here's an example. In debates about the nature of personal identity, those who argue against

117

certain "empiricist" accounts, such as psychological continuity accounts, sometimes use

thought experiments according to which a person undergoes fission – two people are

somehow created from an original person. Richard Swinburne[85], for example, uses such

a thought experiment to show that if some psychological continuity account is correct,

then *both* of the resultant persons should be the same person (by transitivity of identity),

which is supposed to be absurd – no one can be "of two minds" in such a literal sense. If

the psychological continuity theorist comes back and says that we can modestly revise

the account of personal identity by adding that no people with equal claim to

psychological continuity can be the same as the original person, Swinburne will reply that

the thought experiment leads us to see a deeper absurdity with the theory, since it shows

us that it implies that my being identical to some earlier person depends on who else

exists! Now, as a last resort, the proponent of the psychological continuity theory can say

that such thought experiments are bizarre, and so probably metaphysically impossible.

But – and this is the crucial point – *it doesn't matter* whether the scenario is

metaphysically possible; possible or not, it succeeds in getting us to see that such theories

of personal identity are inadequate, since it reveals that they make identity depend on the

wrong sort of thing (in this case, on who else exists).

To recap: thought experiments can be divided up in terms of whether they are

high-flying or *low-flying*. Low-flying thought experiments can be grounded in our

knowledge of the actual world in the ways outlined in our observation-based account;

---

[85] In Shoemaker, Sydney and Swinburne, Richard. *Personal Identity* (Oxford: Basil Blackwell, 1984), pp. 12-19.

high-flying thought experiments cannot be so grounded. Thought experiments can also be divided up in terms of whether they are *modal* or *non-modal*. Modal thought experiments depend for their force upon whether they support a modal claim; non-modal thought experiments do not. With these two distinctions, we can further divide thought experiments into four categories:

	High-flying	Low-flying
Modal		
Non-modal		

TABLE 6.1: Varieties of Thought Experiment

Thus, there is at least logical space for four types of thought experiment: high-flying modal thought experiments, low-flying modal thought experiments, high-flying non-modal thought experiments, and low-flying non-modal thought experiments. Given this fourfold distinction, we are in a better position to distinguish good from bad thought experiments in general, and to utilize the results of chapters 1-3 in particular. To see this, recall that in the first three chapters, we developed and defended a case for Remoteness Modal Skepticism. But according to that view, we don't have justified beliefs with respect to modal claims that can't be grounded in our knowledge of the actual world. Now modal claims that can't be so grounded correspond to those involved in high-flying thought experiments, and the modal claims that can be so grounded correspond to those involved in low-flying thought experiments.

119

Thus, if the arguments of chapters 1-3 are on track, then we can use the current distinction to distinguish good from bad *modal* thought experiments. For since high-flying thought experiments involve claims that can't be grounded in our knowledge of the actual world, they fall prey to the sorts of problems raised in chapters 1-3, and are thus unjustified. By contrast, low-flying modal thought experiments can be grounded in our knowledge of the actual world, and are therefore justified. Thus, given the modal/non-modal thought experiment distinction and the sorts of reasons summarized here, we have a principled basis for distinguishing between good and bad modal thought experiments, and to restrict the use of modal thought experiments to those of the low-flying variety.

Our fourfold distinction also helps us distinguish between good and bad *high-flying* thought experiments. For while our discussion of Remoteness Modal Skepticism in chapters 1-3 raised problems for high-flying *modal* thought experiments, our discussion of *non-modal* thought experiments in chapter 5 suggested that high-flying non-modal thought experiments may well have a significant role to play in Armchair Philosophy, even if they cannot be grounded in our knowledge of the actual world, and thus cannot be known to be metaphysically possible. Thus, we have a way of retaining the utility of high-flying thought experiments even in the face of our case for Remoteness Modal Skepticism. For if high-flying *non-modal* thought experiments can justify significant philosophical claims without appeal to possible worlds, then we can grant the case for Remoteness Modal Skepticism with little negative impact on Armchair Philosophy.

120

6.2 Kinds of Claim, Grades of Strength, Standards of Evaluation & Defeat

We have now completed our discussion of the first set of distinctions relevant to our framework, viz., those pertaining to the variety of thought experiments. The second set of distinctions pertains to certain kinds of philosophical claims: (i) material conditionals and bi-conditionals, (ii) strict conditionals and bi-conditionals, (iii) Roy-style relevant-logic conditionals and bi-conditionals, and (iv) constituent claims and analyses. Each type of philosophical claim has a unique set of truth-conditions, evaluation-conditions, and characteristic proper methods of evaluation and defeat. Let's look at each kind of claim in turn.

Material conditionals and bi-conditionals

Material conditionals and bi-conditionals are (or should be) the least common sorts of significant philosophical claims. For they deal with what is merely <u>actually</u> the case – as opposed to what must or must not be the case – and are thus contingent empirical claims. A material conditional is true just in case the consequent is true whenever the antecedent is true in the <u>actual</u> world. They can thus be defeated by pointing to an <u>actual</u> case that makes the antecedent true and the consequent false. A material bi-conditional is true just in case the conditions expressed on each side of the biconditional sign always march in lockstep together in the actual world. They can thus be defeated by pointing to an actual case in which those conditions come apart. The most frequently relevant means of evaluating conditionals and bi-conditionals is via observation and observation-sensitive theory. As such, thought experiments are often irrelevant for their evaluation. However, it's helpful to discuss them here, since philosophical claims are often stated in terms of

121

material conditionals, and it's important to see the problems with doing so in the context of doing Armchair Philosophy. It's also helpful to discuss them here as a means of contrasting them with the sorts of philosophical claims to follow.

Strict conditionals and bi-conditionals

Strict conditionals and bi-conditionals are nowadays perhaps the most common ways of construing significant philosophical claims. They deal with what is <u>necessarily</u> the case. Thus, unlike material conditionals and bi-conditionals, other worlds besides the actual world are relevant to their truth-conditions, evaluation-conditions, and methods of evaluation and defeat. An interesting feature of strict conditionals and biconditionals is that they assert that, necessarily, whenever one property is present, so is another one – to speak metaphorically, they assert that two or more properties are "companions". We may therefore call them *companion claims*.

A strict conditional is true just in case, <u>in all possible worlds</u>, the consequent is true whenever the antecedent is true. They can thus be defeated by appeal to a <u>possible</u> case that makes the antecedent true and the consequent false. A necessary biconditional is true just in case, <u>in all possible worlds</u>, the conditions expressed on each side of the biconditional sign march in lockstep together. Thus, they can be defeated by appeal to a <u>possible</u> instance according to which the conditions come apart. The most frequently relevant means of evaluating conditionals and bi-conditionals is thus a modal thought experiment. However, given the case for Remoteness Modal Skepticism, only low-flyers are of use in their evaluation. Unfortunately, since many philosophically interesting strict conditionals and biconditionals require evaluation by means of high-flying modal thought

122

experiments, this means that many philosophically interesting strict conditionals and biconditionals fall outside the proper bounds of Armchair Philosophy. Happily, though, a number of philosophically interesting claims originally stated in terms of strict conditionals can be re-construed in terms of claims that are properly evaluable from the armchair. This brings us to our next sort of claim.

Roy's relevant-logic conditionals

Related to the previous sort of philosophical claim are Roy's relevance logic conditionals. As we have seen, these sorts of claims are stronger than strict conditionals, as the truth-conditions of the former, but not the latter, require more than the satisfaction of modal constraints. Thus, as we've seen, P entails Q just in case, for any closed property *being W, being W* has *being Q* as a constituent whenever it has *being P* as a constituent. And we saw that his amounts to saying that P entails Q just in case *being Q* is a *sub-complex* of *being P*. As such, they can be defeated by appeal to at least one closed *being W* such that *being Q* is not a sub-complex of *being P*. Now sometimes – perhaps often – such defeating-conditions obtain at possible worlds in which P exists without Q; accordingly, modal thought experiments may often be relevant. However, as we saw in chapter 5, the truth-conditions of Roy's relevance conditionals don't make crucial reference to possible worlds. Furthermore, there are cases of such defeating conditions that aren't – or at least aren't clearly – cases involving a possible world at which P exists without Q (e.g., there can be cases where P and Q are necessarily co-extensive, and yet Q is not constitutive of P). Finally, there are no doubt cases where there is a possible world at which P exists without Q, but the verification of any such

123

world would require a high-flying modal thought experiment, in which case defeat by means of a modal thought experiment is beyond our reach. Given that these things are so, it can be appropriate in principle to evaluate them by means of *non-modal* thought experiments.

Constituent claims and analyses

The final major kind of philosophical claim is one that says that at least one entity is a constituent, or set of constituents, of another entity. Arguably, this used to be the most common sort of claim defended and criticized by philosophers. Such claims have to do with *essential properties*. Now we've seen that the standard analysis of essential properties construes them modally. In fact, such analyses have often been construed in terms of strict conditionals and biconditionals; as such, the methods of evaluation and defeat for the latter apply to the former. However, I would here like to discuss the truth- and evaluation-conditions for the sort of non-modal account of essential property sketched in chapter 5. Now recall that the latter aren't claims about contingent properties or extrinsic, Cambridge properties; nor are they claims about mere *companion* properties. Rather, they're claims about what properties *enter into the very being* of the thing being analyzed. They therefore go beyond the companion claims of strict conditionals and biconditionals, and claim that one or more properties are *constituents* of some other entity. We may therefore contrast them with companion claims by calling them *constituent claims*. The most cautious sort of constituent claim asserts that one or more entities are constituents of another entity: "*F* is (or is not) essential to *A*." Like strict conditionals, constituent claims asserting that *P* is essential to *Q* are true just in case, in

124

all possible worlds, Q is true whenever P is true. However, they go beyond strict conditionals by adding to this the further truth-condition that Q is a constituent of P.

But the holy-grail type of constituent-claim in philosophy is the *analysis*. An analysis attempts to state the full essence – the recipe or blueprint – of some entity type or token, as in "P $=_{df}$ QRS". Thus, an analysis claim that P $=_{df}$ QRS is true just in case, (i) in all possible worlds, PQRS march in lockstep together, and (iii) QRS are the exact *constituents* of *P*. Given our previous discussion of our non-modal account of essence in terms of Roy's relevance-logic conditionals, we see that the truth-conditions of constituent claims and analyses can, arguably, be reduced to those for Roy's relevance-logic conditionals and biconditionals. As such, the same sorts of thought experiments useful for the latter are also useful for evaluating constituent claims and analyses.

6.3 The Resultant Framework: Thought Experiments and Their Proper Uses

In light of our discussion of various kinds of thought experiments and philosophical claims, we are now in a position to bring together our results to construct an account of the utility of thought experiments. Thus, we have focused on four sorts of claim: (i) material conditionals and bi-conditionals, (ii) strict conditionals and bi-conditionals, (iii) Roy's relevance-logic conditionals, and (iv) constituent claims and analyses. They need to be kept separate, as they have different truth-conditions and standards of evaluation and defeat. Claims of type (i) are contingent claims about the actual world, go true or false solely in virtue of what goes on in the actual world, and are thus to be evaluated or defeated by means of observation and observation-sensitive theory. Claims of type (ii)

125

are claims about other possible worlds, (often) go true of false in virtue of what goes on in other possible worlds, and thus are to be evaluated or defeated by means of modal thought experiments. And claims of type (iii) and (iv) are claims about the essential constituents of a thing, and go true or false in virtue of whether each claimed constituent is essential – enters into the very being of – the entity being analyzed. And since such claims need not automatically make reference to what goes on in other possible worlds, the door is open for asking whether *non-modal* thought experiments, in addition to modal thought experiments, are relevant to their evaluation or defeat. Our account of the variety and utility of thought experiments is summarized in the following table:

	High-flying	Low-flying
Modal	Unjustified; so not useful	Evaluating: -Strict conditionals and bi-conditionals -Roy-style relevance logic conditionals and bi-conditionals -Constituent claims and analyses
Non-modal	Evaluating: -Roy-style relevance logic conditionals and bi-conditionals -Constituent claims and analyses -Counterpossibles (for *per impossibile* cases)	Evaluating: -Strict conditionals and bi-conditionals -Roy-style relevance logic conditionals and bi-conditionals -Constituent claims and analyses

TABLE 6.2: Thought Experiments and Their Proper Uses

We have now sketched our framework, which is an account of the variety and utility of thought experiments. In the next two sections, I continue the task of explicating and motivating the framework in general, and non-modal thought experiments in particular. In the first, I provide a snapshot of a sample philosophical debate at the point where it

126

leads up to and reaches an impasse. I then show how the framework can be used to shed light on the nature and causes of such impasses. In the second part, I show how my framework can be utilized to make progress and push past them.

6.4 Conflated Claims & Illegitimate Slides in Philosophical Debates

Not infrequently in philosophical debate, a thought experiment is offered to defeat a philosophical claim -- e.g., "I can imagine a scenario according to which utility is maximized by raping an innocent child; so utilitarianism can't be true." Then in reply, the defender of the claim says the thought experiment is too bizarre, so probably not a possible one, and thus doesn't justify the modal claim that's supposed to function as the defeater – e.g., "That's a pretty far-fetched case. Furthermore, if we look at the wider and longer-range consequences of the action, we'll surely find that raping the innocent child wouldn't maximize utility." Now in many cases, two sorts of things happen at this point of such debates. In cases of this sort, the critic believes her thought experiment is genuinely metaphysically possible, and thus defeats the claim; the proponent of the claim thinks the thought experiment is probably metaphysically impossible, and thus continues to accept his current view. In cases of this sort, the opponent continues from where we left off at the impasse in the first scenario, saying that *it doesn't matter* if the thought experiment represents a genuine possibility – it still achieves its purpose. Often, and understandably, the defender of the claim is confused as to what the opponent could mean by this reply: if the thought experiment doesn't plausibly depict a genuine metaphysical possibility, how could it defeat the philosophical claim? Since it's not clear

127

what to make of such disputes, their effect isn't much different than that of straight impasse cases.

Impasses: Diagnoses

What's going on in these two sorts of scenario? With our framework before us, we're able to give a diagnosis. In cases of the first sort, the philosophers agree on the sort of claim advanced – a strict conditional or biconditional – but disagree about the force of the thought experiment. In our utilitarian vs. non-utilitarian case, both construe the claim in dispute as a necessary biconditional, and thus to be true just in case, *in all possible worlds*, moral rightness and maximized utility march in lockstep. Thus, they both agree that a *modal* thought experiment is the relevant sort. Unfortunately, the opponent's thought experiment is of the high-flying sort (e.g., the non-utilitarian's "utility-maximizing child rape" case). Thus, while the opponent may find it convincing, it has little effect on the defender.

In the second sort of case, the philosophers are most likely talking past each other. That is, (perhaps) without either opponent knowing it, *they're debating different construals of the same basic claim.* The opponent is engaging in dialectical moves appropriate when construing the claim as a *constituent claim* or an *analysis,* and is thus applying the appropriate standards, and methods of evaluation and defeat, of the claim so construed – i.e., a non-modal thought experiment aimed at showing that utility maximization isn't a *constituent* of moral rightness. But unfortunately, the defendant is engaging in dialectical moves appropriate when construing the claim as a *strict conditional* or *biconditional* – a *companion* claim -- and is thus applying different

128

standards of evaluation. To be more specific, he's taking his opponent's thought experiment to be aimed at justifying a metaphysically possible world at which moral rightness and maximal utility come apart, and is thus using a *modal* thought experiment. The result is that the two philosophers are talking past each other, and one or both feel unsure about whether progress has been made in their debate.

Impasses: cures

The cure depends on which of the two sorts of cases we're talking about, but both cures stem from our framework. The first scenario – agreement on the construal of the claim, but disagreement about the force of the modal thought experiment – is susceptible to progress by taking the following steps:

(i) Admit the impasse about the strict conditional or biconditional construal of the core claim, due to the controversial nature of the high-flying modal thought experiment.

(ii) Then *move on* – i.e., switch to a constituent claim or analysis construal of the core claim. In our utilitarian vs. non-utilitarian case, this would amount to switching fro

    1. $\Box$(An act x is morally right iff x maximizes utility)

to

    1*. An act x is morally right $=_{df}$ x maximizes utility.

(iii) Switch to the corresponding appropriate standards and methods of evaluation and defeat for constituent claims and analyses. In our utilitarian vs. non-utilitarian case, they should switch from looking for a defeater of the strict biconditional (1) -- i.e., for a

129

thought experiment that shows a possible world at which moral rightness and maximized utility come apart (aren't *companions*) -- to looking for a defeater of the *analysis* (1*) – i.e., looking for a thought experiment that shows that maximized utility doesn't "enter into the very being of" (isn't a *constituent* of) moral rightness, irrespective of whether they're necessarily co-extensive.

A key moral here is that when you hit an impasse on whether a high-flying modal thought experiment defeats a *companion* claim – a necessary conditional or biconditional -- you should "shift gears" by re-construing it as a *constituent* claim or analysis, and then evaluate it via a *non-modal* thought experiment.

The second sort of case requires a similar cure, with one addition: Both disputants must first come to see that there are different ways to construe the philosophical claim in question, and to decide which one they'll discuss. In particular, they need to get on the same page about the fact companion claims are one sort of claim, and constituent claims are another, each with its own unique set of truth-conditions, evaluation-conditions, and characteristic proper methods of evaluation and defeat. Once they do this, they need to reach agreement on which construal they'll discuss, and then make the appropriate changes and calibrate their evaluation standards and methods, and go from there. If they agree to construe the basic claim as a constituent claim or analysis, then they may well make progress via a non-modal thought experiment. If, on the other hand, they agree to construe the basic claim as a strict biconditional and hit an impasse, then their case turns into an instance of the first Scenario. As such, they should go through the steps (i)-(iii) that constitute the cure for cases of that sort.

130

Illustrating the diagnosis

To make these points more concrete, let's illustrate them with two versions of a debate

between a utilitarian and a non-utilitarian:

Steve: Moral rightness should be analyzed solely in terms of maximized utility.

Betty: That can't be right, for consider this thought experiment:

> On Alpha Centauri, people are so constituted that they achieve the highest
> level of happiness and satisfaction – say, on a par with wining the Nobel
> Peace Prize – when they rape little children from another race. Running
> out of children of other races on their planet and nearby planets, they
> transport to Alpha Centauri, in Star Trek fashion, some of the children
> from Earth, and use their Men in Black-style zappers to erase our
> memories of the children. Perhaps they zap the minds of the children with
> it as well, and perhaps also fiddle with their psychology and/or physiology
> so that they feel merely *melancholy* about being raped. Soon after,
> maximized utility is restored on Alpha Centauri.

In the scenario, then, it is *prima facie* plausible that such an act would maximize utility,

yet it's *prima facie* implausible that the act is morally right. Thus, since such a scenario

is metaphysically possible, we see that moral rightness shouldn't be construed in terms of

maximized utility.

Steve: You're case is wacko. It's too bizarre to carry conviction as a genuine possibility.

In any case, if we expanded our view of the negative consequences of raping the innocent

children, we'd see that doing so would not maximize utility. Thus, your crazy scenario

does not give good reason to think that the properties of moral rightness and utility

131

maximization come apart in some possible world. So, the scenario is not a genuine counterexample.

Betty: I think your response is clearly *ad hoc*. But in any case, even if the scenario isn't metaphysically possible, it doesn't matter. For it *still* shows that *your account makes moral rightness depend on the wrong sort of thing*, since it got us to see that the moral status of the act doesn't depend on whether it turns out that the act does or doesn't maximize utility.

Steve: Huh? What can you possibly mean by that? If your scenario isn't genuinely metaphysically possible, then in what sense is it a telling case against my account of moral rightness?

Betty: It just is. Look, I don't know how I can get my point across in a way that's plainer than by pointing to my scenario. If you can't just *see* that it shows that you're view is implausible, I don't know what else to say. I guess we just have different intuitions, and those whose intuitions are like mine will find the case convincing; those with intuitions like yours will not.

What's going on in this debate, and why does disagreement remain? The first thing to notice is that Steve and Betty are conflating analyses with necessary bi-conditionals.

132

Steve originally gave an *analysis* of moral rightness in terms of utility maximization – the one expressed by our (1*) above:

1*. An act x is morally right $=_{df.}$ x maximizes utility.

But, based on the way Betty argues later, she's treating Steve's claim as though it were a *necessary biconditional* – our (1) above:

1. $\square$(An act x is morally right iff x maximizes utility)

This leads to a conflation of corresponding (a) truth-conditions and (b) standards and methods of evaluation and defeat. Thus, intending to attack (1) – not (1*) -- Betty argues as follows:

1) I can conceive of a scenario according to which an act is utility-maximizing yet morally wrong.

2) If I can conceive of a scenario according to which an act is utility-maximizing yet morally wrong, then there is a possible world at which there is a case of utility-maximization without moral rightness.

3) If there is a possible world at which there is a case of utility-maximization without moral rightness, then $\sim\square$(x is morally right iff x maximizes utility).

Therefore,

4) $\sim\square$(x is morally right iff x maximizes utility).

The particular scenario she uses is both high-flying and modal; thus, as you might guess, Betty and Steve are at an impasse with respect to the truth of (1). Interestingly, though, Betty claims that *it doesn't matter*. However, since she never explicitly changes her construal of the claim from a necessary bi-conditional to an analysis, she can't explain why this matters in a way that would sway Steve. Thus, they can't resolve the dispute

133

unless one can convince the other about the possibility of the scenario depicted in the thought experiment.

Illustrating the cure

Now for the second payoff. We've already seen one payoff of our framework in terms of diagnosing the problems in the debate, and the cause of the stalemate. The second payoff will consist in curing them – i.e., in utilizing our framework to avoid their pitfalls and make progress in their dispute. In light of our framework, we can reformulate the debate between Betty and Steve in a way that helps them to avoid conflating types of claims, keep track of the use of each type of claim, and thus employ the most relevant and effective sorts of thought experiments. Let's illustrate how this works out in practice:

Steve: Moral rightness is properly analyzed in terms of maximized utility: an act is morally right iff it maximizes utility.

Betty: It's not clear exactly what sort of claim you're making. At first, it sounded as though you were going to offer an *analysis* of the property of moral rightness in terms of utility maximization. But then you re-stated the claim as an ordinary *biconditional*. Now, based on the first thing you said, I don't think you meant to switch gears and just claim that, *as a matter of fact*, moral rightness and utility maximization march in lockstep through the *actual* world. That's a contingent empirical claim, and isn't properly addressed by Armchair Philosophers like you and I. On the other hand, I think you meant your biconditional to be taken as a *strict* biconditional: necessarily, something is morally

134

right iff it maximizes utility. Now I grant that justifying strict biconditionals shows that the presence of one property is a reliable indicator of the presence of another property. I even grant that such biconditionals can sometimes justify that the two properties are identical. Nevertheless, analyses aren't the same thing as mere strict biconditionals. Thus, to correctly express your original intention, we should construe your claim about moral rightness as follows:

1*. x is morally right $=_{df}$ x maximizes utility

On this construal, then, you're not just saying that the two properties march in lockstep merely in the actual world. Nor are you just saying that they do so across all possible worlds – as *companions*. Rather, you're going beyond this and saying that the former property *just is* the latter property – it's the sole *constituent*. Have I stated your intended claim correctly?

Steve: Well, that was a mouthful! But once you put it like that, I'm not sure which claim I'd like to make. I guess I'd say that I accept each one with varying degrees of conviction: I definitely accept the biconditional construal, since it follows from the necessary bi-conditional construal that I also accept. However, I'm not as convinced about the analysis construal.

Betty: Ok, that puts me in a better position to discuss the issue with you in a way that prevents needless disputes caused by misunderstandings about what is being asserted. Given your varying degrees of conviction on the varying construals, I see that we're not

135

that far apart in our views. For I don't find the first claim egregiously implausible, although I'm not completely comfortable with the strict bi-conditional construal. However, my intuitions go strongly against the analysis construal. Let's start our discussion, then, with the strict bi-conditional:

1. $\Box$(x is morally right iff x maximizes utility)

Since it's a strict bi-conditional, it's true just in case there is no possible world at which the two properties come apart. Thus, to refute it, I'd have to give good reason to think that there's at least one possible world at which they come apart. Well, I can't make a quick argument from actuality, since I don't know of an *actual* case in which they come apart. Thus, I need to come up with a *modal* thought experiment that would justify this claim. Aha! How about this?

> On Alpha Centauri, people are so constituted that they achieve the highest level of happiness and satisfaction – say, on a par with wining the Nobel Peace Prize – when they rape little children from another race. Running out of children of other races on their planet and nearby planets, they transport to Alpha Centauri, in Star Trek fashion, some of the children of Earth, and use their Men in Black-style zappers to erase our memories of the children. Perhaps they zap the minds of the children with it as well, and perhaps also fiddle with their psychology and/or physiology so that they feel merely *melancholy* about being raped. Soon after, maximized utility is restored on Alpha Centauri.

Thus, we are *prima facie* justified in thinking that such a scenario is metaphysically possible. And if so, then the necessary biconditional is defeated. Granted, it's a relatively *high-flying* modal thought experiment, but not so disanalogous to actual experience as to render it implausible.

136

Steve: That's stretching it a bit, don't you think? After all, that's a pretty bizarre case. It's certainly too bizarre to carry conviction as a genuine possibility. In any case, if we expanded our view of the negative consequences of raping the innocent children, we'd see that doing so would not maximize utility. Thus, your crazy scenario does not give good reason to think that the properties of moral rightness and utility-maximization come apart in some possible world. So, the scenario is not a genuine counterexample.

Betty: Really, you find it unconvincing? It seems to me rather that your resistance of the thought experiment is *ad hoc*. But I can't think of a low-flying modal thought experiment that could help us make progress in evaluating the strict biconditional construal of your claim. Perhaps we should agree to disagree on this construal, and see if we can make progress on the analysis construal.

Steve: I think you're right. Let's discuss the analysis construal – our (1*).

Betty: Ok. Well, the truth-conditions are different for analyses, as we've seen. Thus, if I can show that a candidate constituent of the analysis is the wrong sort of property to be a constituent of -- "enter the very being of" -- moral rightness, or show that the analysis leaves out something crucial, that would be enough to defeat the analysis – even if moral rightness and maximized utility never come apart. So, for example, if I could show that, necessarily co-extensive or not, maximized utility is somehow *unacceptably extrinsic to* moral rightness (e.g., in a way similar to how being a truth-tracker is necessarily co-

137

extensive with yet unacceptably extrinsic to Cartesian clear and distinct perceptions), or

that it makes moral rightness depend on the wrong sort of thing, then that would show

that the property is a *companion* of moral rightness at best, and not a *constituent* of it. If

so, then it would show that your analysis is false – again, *even if moral rightness and*

utility-maximization are necessarily coextensive. Thus, a <u>non-modal</u> thought experiment

– even a <u>high-flying</u> non-modal thought experiment -- could settle our dispute in

principle. For then it's *irrelevant* whether it depicts a genuine metaphysical possibility.

Steve: I hadn't thought of it that way before, but that sounds right. But first let me see if I

have your reasoning right. Analyses are claims about constituents; strict biconditionals

are about companions. To defeat companion claims, you need to point to a possible

world at which the properties come apart. But since analyses are constituent claims, you

can do something that's easier in principle: point out that, companion or not, the property

at issue isn't a constituent. Conceivably, this can be done in a number of ways, but one

way is to find a non-modal thought experiment that enables you to see that a candidate

constituent of the analysis is unacceptably *extrinsic* to the analysandum, or perhaps helps

you to see that the analysis makes the *analysandum* depend on the wrong sort of thing, or

some such insight into the relationships between the properties involved.

Betty: I couldn't have said it better myself. So now that we've gotten clear on these

matters, how about this: take my original high-flying modal thought experiment, and

slightly modify it so as to make it a high-flying *non-modal* thought experiment. Thus,

138

consider the Star Trek-like scenario about the children and the raping Alpha Centaurians, but *grant* that the act may not be possible. Still, isn't it clear that that thought experiment reveals that the analysis *makes moral rightness depend on the wrong sort of thing?* After all, our judging it to be wrong that the children are raped is *independent* of whether or not it turns out, on closer inspection, that it wouldn't maximize utility. That is, even if both properties never come apart, utility maximization isn't what's doing the grounding here: at best it's a mere *companion* property of moral rightness, not a *constituent* of it.

Steve: I wouldn't have believed it, but I think you've moved our discussion forward via armchair methods – viz., thought experiments -- without appeal to genuinely possible worlds. Now, I may not be completely convinced that you're right, but at least your new, non-modal thought experiment pressures me to explain why your moral wrongness intuition seems to be disconnected from the act's utility – I can't just dismiss this thought experiment as I could with its modal counterpart on the grounds that its maximized-utility-without-moral-rightness defeater world is not really possible; *for it doesn't rely on the existence of such a world.*

Lessons

Betty and Steve made progress in their debate. How did they do it? First, they distinguished the various sorts of philosophical claims – material conditionals and biconditionals vs. necessary conditionals and biconditionals vs. constituent claims and analyses. Then, they laid out their respective truth-conditions, and their corresponding

139

conditions of evaluation and defeat. Next, they distinguished the variety of thought experiments, and pointed out which ones were relevant to each sort of claim.

With these points and distinctions laid out, they were then able to make progress without getting stalled due to confusions over claims, their truth- and evaluation-conditions, and the stalemates that can ensue. Thus, when construed as a *strict biconditional* -- as a *companion* claim -- they evaluated it by using a high-flying *modal* thought experiment. But when the thought experiment couldn't settle their disagreement on that construal of the claim, they pushed forward by construing the claim as an *analysis* – as a *constituent* claim -- and thus evaluated it by switching to a high-flying *non-modal* thought experiment. And this, in turn, yielded progress in the debate.

6.5 How to Defeat Modal Claims Without Appeal to Possible Worlds

Before moving on to sketch other sample applications of our framework, I'd like to make one last suggestion about its utility for doing Armchair Philosophy. It's not uncommon for philosophers to argue that if they can justify a necessary equivalence between two sets of properties, then this is good evidence that one is to be analyzed in terms of the other. If so, then it appears that they accept the following conditional:

> (NEA) If we're justified in taking two sets of properties to be necessarily
>
> coextensive, then we're justified in taking one as an analysis of the other.

But if so, then if a high-flying non-modal thought experiment undermines an analysis in a way that's relatively uncontroversial, then we can reason in G.E. Moore Shift, *modus tollens* fashion that the two sets of properties probably *aren't* necessarily co-extensive.

140

So, for example, consider our earlier example of psychological continuity accounts of personal identity. We can construe the claim (generically, for our purposes) in the form of a necessary biconditional:

1. Necessarily, person A is the same person as B iff B is psychologically continuous with B.
2. If we're justified in taking two sets of properties are necessarily coextensive, then we're justified in taking the one as an analysis of the other.
3. We're not justified in taking psychological continuity as an analysis of personal identity.
4. Therefore, we're not justified in taking psychological continuity and personal identity as necessarily coextensive.

The key premise is (3). Now of course, one could try to justify it by means of a modal thought experiment; that is, by arguing that there is a possible world at which personal identity and psychological continuity come apart. This seems to be what Swinburne took himself to be doing with his story about personal fission. However, this story is clearly a high-flying thought experiment – it's remote from ordinary experience. As such, you can bet that philosophers will question (and not without reason) whether such a story is really metaphysically possible. Philosophers may then want to criticize the thought experiment, thus leading to a stalemate. As we all know, this happens all too often in philosophical debate – think about debates regarding Frankfurt-examples, utilitarianism vs. deontology, etc.

However, suppose we construe the thought experiment non-modally. Thus, suppose one argues that *whether or not* personal fission is metaphysically possible, it *still* shows us that the psychological continuity account makes personal identity depend on the wrong sort of thing -- viz., on whom else exists – in which case psychological continuity

141

is *extrinsic* to personal identity. If so, then we can undercut or rebut (3), and thereby undercut or rebut the psychological account of personal identity, in a way that circumvents the sort of stalemate caused by doubts about the metaphysical possibility of the story.

To come at the point from another direction: recall our critique of modal accounts of essential properties in chapter 5. One lesson from that discussion was that necessary coextension is necessary but not sufficient for essential propertyhood: B may be necessarily coextensive with A, but for all that, B may not be essential to A. This is because there are at least two conditions for B to be essential to A:

(i) B is necessarily coextensive with A.

(ii) B is a constituent of – enters into the being of – A.

Now modal thought experiments seek to undermine analyses or constituent claims by attacking clause (i), and this is typically done by means of a modal thought experiment. However, such thought experiments often fail to carry conviction, especially when they're high-flyers. Now what I'm suggesting is that there is another way to attack essence or constituent claims, viz., by attacking clause (ii). And one need not do this by means of a modal thought experiment; rather, one can do so by construing the relevant thought experiment non-modally. If so, then if one also accepts NEA, then one can undermine the modal claim – the strict biconditional – without appealing to possible worlds. In this way, then, it's possible to defeat a modal claim without appeal to possible worlds and modal thought experiments.

142

We have now laid out our framework for that fragment of Armchair Philosophy for which the use of thought experiments is germane. We have also shown how it can be used to make progress and push past impasses in principle. Finally, we have demonstrated how to use the framework in a specific case – i.e., in one for which only a high-flying non-modal thought experiment may be of use. From here, we can see how to use the framework in the other sorts of cases. Thus, consider an example for which a modal thought experiment is useful: the Traditional Analysis of Knowledge (TAK). We can construe claims germane to TAK as either constituent claims/analyses, strict conditionals/bi-conditionals, or as Roy-style conditionals/bi-conditionals. For as discussed in chapter 2, we have good low-flying modal thought experiments to evaluate it, such as the Nogot/Havit case, and thought experiments of this sort are sufficient to defeat all three sorts of claims here.

Finally, we see that if there are any cases for which we can find neither a good low-flying modal thought experiment nor a good non-modal thought experiment, our framework leads us to deem the corresponding claim beyond our ability to evaluate it – whether it be construed as a strict conditional, a Roy-conditional, or a constituent claim -- at least by means of a thought experiment. I'm not foolish enough to say that my framework rules out the possibility of such cases, although my hope is of course that they are very few. Rather, my claim is the humbler one that the scope of the utility of my account of thought experiments goes at least as far as that of any other account, and given the force of the case for Remoteness Modal Skepticism, a good deal farther.

143

I've suggested several examples of non-modal thought experiments: the Twin Earth Case, the Personal Fission Case, and the Clear and Distinct Perceptions Case. We have also seen, in some detail, how to use such thought experiments to evaluate claims in our discussion of the Utilitarianism Case. However, given that non-modal thought experiments are perhaps unfamiliar to us, I will close by looking at two more non-modal thought experiments in a bit more detail. This will also contribute to illustrating the utility of the framework just sketched.

6.6 Sample Application I: The Principle of Alternate Possibilities[86]

Thus, consider the debate about Frankfurt-Counterexamples.

Setup

It is commonly thought that moral responsibility requires the ability to do otherwise. Harry Frankfurt dubbed this principle the Principle of Alternate Possibilities (PAP). He gave a thought experiment that seemed decisive against it:

> Suppose someone – Black let us say – wants $Jones_4$ to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until $Jones_4$ is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that $Jones_4$ is going to decide to do something *other* than he wants him to do. If it does become clear that $Jones_4$ is going to decide to do something else, Black takes effective steps to ensure that $Jones_4$ decides to do, and that he does do, what he wants him to do. Whatever $Jones_4$'s initial preferences and inclinations, then, Black will have his way...Now suppose that Black never has to show his hand because $Jones_4$, for reasons of his own, decides

---

[86] The following is a summary of the relevant points in Leon, Felipe and Tognazzini, Neal. "Why Frankfurt-Examples Don't Need to Succeed to Succeed", *Philosophy and Phenomenological Research* (forthcoming).

144

to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones₄ will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it.[87]

So in this scenario, Jones is morally responsible for his actions, and yet he couldn't have done otherwise (since Black would have intervened to ensure that he does it if he were about to choose something different). But if this is right, then it appears that PAP is false.

A number of philosophers have argued that neither this original version of a Frankfurt-example nor any subsequent version succeeds in defeating PAP. For they argue that the agent in the scenario is still able to do *something*, viz., show a *sign* that he intends to act in a certain way; if so, then such thought experiments fail to defeat PAP.

The current state of the debate: stalemate

Thus, the debate about Frankfurt-examples goes back and forth in this manner, with those on one side of the debate generating more and more refined versions of Frankfurt-examples, and those on the other side generating doubts about whether these really involve a morally responsible agent who's unable to do otherwise. From all appearances, the debate shows little signs of waning. One explanation for this naturally suggests itself: the interlocutors in the debate have reached an impasse.

Whether I'm right about this, I think our framework can provide a way to make progress in the debate about Frankfurt-examples. I of course don't mean to suggest anything as ambitious as that the framework provides a way to unequivocally resolve all

---

[87] Frankfurt, Harry. "Alternative Possibilities and Moral Responsibility", *Journal of Philosophy* 66 (1969), p. 835ff.

145

philosophical disputes; rather, I intend to make the weaker claim that the framework can be used as one more helpful tool to push debates forward.

Diagnosis

Where do we start in the task of pushing past the apparent impasse? As before, it is helpful to first get clear on the claim under dispute. As mentioned above, the main point of contention is PAP, which claims that moral responsibility requires the ability to do otherwise. But there are at least three ways to construe PAP, each one corresponding to a conditional of the sort discussed in the previous two chapters: as a material conditional, as a strict conditional, and as a constituent, Roy-conditional. Thus, we can state the three construals of PAP as follows:

> (PAP1): An agent is morally responsible for an action only if he could've done otherwise.

> (PAP2): Necessarily, an agent is morally responsible for an action only if he could've done otherwise.

> (PAP3): It is constitutive of an agent's moral responsibility for an action that he could've done otherwise.

Corresponding to each construal of PAP is a corresponding set of truth-conditions, evaluation conditions, and methods of evaluation and defeat. Thus, PAP1 is true iff there is no <u>actual</u> case in which an agent is morally responsible for an action for which they couldn't have done otherwise; PAP2 is true iff there is no <u>possible</u> case in which an agent is morally responsible for an action for which they couldn't have done otherwise; and PAP3 is true iff (i) there is no possible case in which an agent is morally responsible for an action for which they couldn't have done otherwise, and (ii) the ability to do otherwise is <u>constitutive</u> of moral responsibility.

146

Now although Frankfurt originally formulated PAP as a material conditional, i.e., as PAP1, it would be uncharitable to say that he intended it as such. For that would imply that his original version of the counterexample to PAP has no bearing on its truth-value. For recall that he merely gave a *hypothetical* counterexample, and a material conditional requires an *actual* counterexample of the latter.

What about PAP2? On this interpretation, Frankfurt's original counterexample is indeed relevant. For then his hypothetical case constitutes a counterexample to the strict conditional here if the former turns out to be metaphysically possible. Furthermore, his case seems to be a low-flying modal thought experiment: it depends for its force upon a modal claim that can be traced back to our knowledge of the actual world. For we know from experience and our folk theory of how the world works that people can be manipulated in various ways to get them to do what we want. Therefore, the modal claim enjoys significant initial justification.

However, defenders of PAP have a not implausible reply when they say that, in cases such as this, the agent can at least do *something* – e.g., they can *make a sign* that they intend to do something. Now of course it could turn out later that a philosopher comes up with an ingenious thought experiment according to which the agent is morally responsible and yet cannot even give a sign of this sort. But if the current state of the debate is any indication, there is little reason to be optimistic.

Cure

But even if I'm wrong about that, I'd like to propose a hitherto unexplored way to make progress in the debate about Frankfurt-examples and PAP, viz., by focusing on an

147

evaluation of PAP3 instead of PAP2. Toward this end, recall that according to PAP3, the

ability to do otherwise is not only necessarily co-extensive with moral responsibility, it's

also part of what *makes* moral responsibility *what it is*. To put it in the language of our

framework, PAP3 states that the ability to do otherwise is a *constituent* of moral

responsibility, and not just a *companion* of it. Thus, PAP3 is stronger than both PAP1 and

PAP2; as such, it's easier to evaluate in principle. Now, given clause (i), the standard

approach to criticizing the Principle of Alternative Possibilities via modal thought

experiments is certainly legitimate. However, clause (ii) allows for an alternative and,

arguably, much easier method of defeat. For it leaves open the possibility of showing that

the ability to do otherwise is not constitutive of moral responsibility *without appeal to*

possible worlds at which the properties come apart. This could be done if one could

show that, whether or not the ability to do otherwise is necessarily co-extensive with

moral responsibility, the former is *irrelevant* to the latter. John Martin Fischer raises a

case that seems to defeat PAP3 in just this way:

> Some months ago, I was at the local supermarket. When I got to the
> checkout counter, the checker asked, "Would you like a paper bag or
> plastic?" I thought for a moment and replied, "Plastic is fine." Then the
> checker smiled and said, "It's a good thing – I see we only have plastic!"[88]

In this case, it seems clear that John is responsible for choosing plastic. It also seems

clear that he couldn't have gotten paper. However, he still could've done something

different – he could've *chosen* paper. Thus, it doesn't give us a possible case (because it

isn't an *actual* case) in which moral responsibility exists without the ability to do

---

[88] Fischer, John Martin. "Responsibility and Alternative Possibilities", in his *My Way:*
Essays on Moral Responsibility (Oxford: Oxford University Press, 2006), p. 62, n. 22.

otherwise. However, it *does* seem to be a case in which the ability to do otherwise is *irrelevant* to moral responsibility. For while John couldn't have gotten paper, that fact didn't play a role in his deliberations for choosing plastic, as he thought that choosing paper was a real option. Thus, whether or not the ability to do otherwise and moral responsibility are *companions*, the case seems to show that the former is not a *constituent* of the latter. And if that's right, then the case defeats PAP3, and it does so without depending on an appeal to other possible worlds at which the two properties come apart.

If the preceding argument is successful, then it defeats PAP3 without necessarily defeating PAP2; that is, it undercuts the claim that the ability to do otherwise is a constituent of moral responsibility, but in a way that doesn't clearly undercut the claim that it's a companion of moral responsibility. However, once this is done, it is open to the proponent of the argument to try and push the debate further by arguing as follows: It's not uncommon among philosophers to argue that if they can justify a necessary equivalence between two sets of properties, then this is good evidence that one is to be analyzed in terms of the other. If so, then it appears that they accept the following conditional:

> (NEA) If two sets of properties are companions, then we're justified in
>
> taking the one as a constituent of the other.

But if so, then if the proposed non-modal thought experiment is good evidence for rejecting the constituent claim, then its proponent can reason in *modus tollens* fashion that that the two sets of properties probably aren't companions, either. In this way, then, one can *indirectly* attack PAP2 via first attacking PAP3, and then appealing to NEA in

149

the way just mentioned. If so, then one can not only evaluate constituent claims without appeal to possible worlds, but one can also evaluate *modal* or companion claims without direct appeal to possible worlds.

6.7 Sample Application II: The Standard Analysis of Essential Properties

Setup

Things have at least some of their properties at every possible world in which they exist. Call this phenomenon 'rigid exemplification'. What is the relationship between essential properties and rigid exemplification? A popular view has it that the latter is correctly *analyzed* in terms of the former – i.e., essential propertyhood *just is* rigid exemplification:

> (EP1) F is an essential property of x $=_{def}$ x has F at all (metaphysically) possible worlds in which x exists.

A slightly weaker account of the relationship has it that while the latter can't be analyzed in terms of the former, the latter and the former are yet *necessarily co-extensive*[89]:

> (EP2) Necessarily, F is an essential property of x iff x has F at all (metaphysically) possible worlds in which x exists.

I think that EP1 and EP2 are both false. That is, I think that although rigid exemplification is surely a *necessary* condition for a property to be essential to its bearer, it isn't sufficient. For I think it's possible for a bearer to rigidly exemplify a *non-essential* property. In this section, I'll use our framework to show how one might argue for this.

---

[89] As, for example, triangularity and trilaterality are distinct yet necessarily co-extensive properties.

A non-modal thought-experiment against standard accounts

As briefly discussed in Chapter 5, Kit Fine has argued against modal analyses of essential properties. The argument we saw there was based on a counterexample involving Socrates and his singleton set. Recall that in that case, Socrates had the property of being the only member of his singleton set in all possible worlds in which he exists, and yet, intuitively, this property isn't one of Socrates' essential properties. One might reply that this is merely one of Socrates' extrinsic properties, and that while it may be hard to define essential properties in such a way as to distinguish genuinely essential from merely extrinsic-but-rigidly-exemplified properties, in practice one can often do so. However, one may still wonder how this matters since, strictly speaking, Fine's case is a genuine counterexample; in particular, it shows that rigid exemplification is not a sufficient condition for essential propertyhood. Still, one might nonetheless worry that there is a way to qualify the standard modal analysis to handle this case, and in a way that retains the spirit of the modal account by taking rigid exemplification to be the central aspect of essential propertyhood.

I'd like to advance a counterexample to the standard, modal analysis of essential properties, as well as the strict equivalence claim. I think that my counterexample is an advance over Fine's, since mine involves a case of a bearer rigidly exemplifying a non-essential *intrinsic* property. The thought experiment can be put as follows:

> The Pinky/Winky Case:
> Suppose there are two Anselmian (or quasi-Anselmian) necessary beings, Pinky and Winky. Now while it's not constitutive of Pinky that he *has* a pink after-image throughout his existence, it *is* constitutive of Pinky that he has an aching *desire* to have one. And as it turns out, it's constitutive of Winky that he *satisfies* all the harmless desires of Pinky (because, say, it's

151

constitutive of Winky that he's a perfectly loving being, and since Pinky is Winky's beloved, he inevitably satisfies all his harmless desires – or at least desires of *this* sort). In this case, then, it seems that Pinky has the property of having a pink after-image in all the worlds in which he exists, and yet the after-image is not essential to Pinky.

According to the Pinky/Winky case, then, Pinky rigidly exemplifies a property he bears intrinsically, viz., a pink after-image. Yet, intuitively, pinkness is not essential to Pinky. Therefore, it appears that EP1 and EP2 are both false.

An objection

I have just argued that we've been wrong about the relationship between rigid exemplification essential propertyhood: the former is neither constitutive nor a companion of the latter. For the Pinky/Winky case seems to show that objects can rigidly exemplify *non-essential* properties. However, one may remain unconvinced. For they may doubt that the Pinky/Winky case is genuinely metaphysically possible, and that if it isn't, then neither EP1 nor EP 2 has been defeated.

Reply

However, our framework provides the materials for a reply. For the objection assumes that the Pinky/Winky case is a *modal* thought experiment. Now if it *were* construed that way, then the objector would certainly be right to say that it has no force against either version of EP. For if it were a modal thought experiment, then the force of the thought experiment would depend upon whether it justifies a modal claim, and we've seen that only low-flying modal thought experiments can do that. But the Pinky/Winky thought experiment is pretty clearly a high-flying thought experiment: we haven't seen Pinky or Winky, let alone Anselmian or quasi-Anselmian beings of any sort; nor do we have the

152

credible testimony of others of the existence of such beings. Thus, we can't generate an argument from types to tokens for their existence; nor can we support claims of their possible existence from our folk theory of how the actual world works; nor are such beings relevantly similar to things we've observed in the actual world, in which case we can't justify their possible existence via an argument from analogy. The Pinky/Winky case is therefore a high-flying experiment, in which case it would have no epistemic force if it were a *modal* thought experiment. So construed, therefore, it is unable to directly defeat either EP1 nor EP2.

However, since the Pinky/Winky case is a high-flying *non-modal* thought experiment, its force is not deflated by its inability to justify the possibility of rigid exemplification without essential propertyhood. For on the latter construal, the thought experiment is directly attacking the claim that rigid exemplification is *constitutive* of essential propertyhood – i.e., EP1 -- and not whether the former is a *companion* of the latter – i.e., EP2. And with respect to the former task, the thought experiment seems to be a success. For the example gets us to see that properties are not essential *in virtue of* being rigidly exemplified by their bearers; rather, the in-virtue-of relation between essentiality and rigid exemplification seems to run in the other direction. And this non-modal intuition puts us in a position to see that rigid exemplification isn't constitutive of essential propertyhood. Therefore, even if the Pinky/Winky case doesn't refute EP2, it nonetheless shows that EP1 is inadequate as an *analysis* of essential propertyhood.[90]

---

[90] One could make similar moves when evaluating Roy-conditionals: the thought experiment enables one to see that the in-virtue-of relation runs from essential propertyhood to rigid exemplification, and not vice-versa; this gives one indirect

153

Now suppose we're right that when construed as a non-modal thought experiment, the

Pinky/Winky case defeats EP1. If so, then, as with our discussion of Frankfurt

counterexamples and PAP, one might go on to use this result as a basis for undercutting

EP2. For again, we can appeal to NEA, and then argue that since rigid exemplification

isn't constitutive of essential propertyhood, doubt is cast on their companionship. And if

that's right, then EP2 is undercut as well.

6.8 Conclusion: Responding to the Challenge of Modal Skepticism

It's fitting to conclude by re-visiting the problems we met at the beginning of this

dissertation, and to see how the results of our inquiry can be brought to bear on their

evaluation. Thus, we started our discussion by considering a formidable threat to classical

Armchair Philosophy, viz. BAAAP:

> **The Basic Argument Against Armchair Philosophy (BAAAP)**
> 1. Armchair philosophy is legitimate only if thought experiments give us insight into the nature of things.
> 2. Thought experiments give us insight into the nature of things only if they justify modal claims.
> 3. Thought experiments don't justify modal claims.
> 4. So, thought experiments don't give us insight into the nature of things.
> 5. So, armchair philosophy is not legitimate.

We've seen that while many adherents of armchair philosophy (in effect) use the Modal

Epistemology Response to respond to BAAAP, we've also seen (in chapters 1-3) that a

case for Remoteness modal Skepticism very probably rules out an unqualified version of

---

evidence that essential propertyhood isn't a sub-complex of rigid exemplification; and
this, in turn, puts one in a position to negatively evaluate the relevant Roy-conditional.

154

such a response. Together, these points generated pressure to either adopt a concessive response to BAAAP, and thereby abandon Armchair Philosophy, or to hold on to Armchair Philosophy and generate a plausible, yet novel, non-concessive response. And while we granted, at least for the sake of argument, the legitimacy of concessive responses, in later chapters we developed various strands of a novel, non-concessive response to BAAAP, which culminated in an account of the variety and utility of thought experiments. Thus, we distinguished between various sorts of thought experiment on the one hand – high-flying, low-flying, modal and non-modal --, and various sorts of philosophical claim, on the other – strict conditionals, Roy-conditionals, and analyses. We then argued that different thought experiments are useful for evaluating different sorts of claim. And finally, we used these points and distinctions to construct a framework for doing a significant fragment of traditional Armchair Philosophy.

With our framework laid out, we now have the materials for a nuanced response to the challenge to Armchair Philosophy found in BAAAP. Thus, we've seen that while our framework is committed to (1), it does give us a basis for principled criticisms of (2) and (3). Taking these in reverse order, the plausibility of (3) is based on a failure to appreciate that we have a solid, non-mysterious account of when a thought experiment can justify a possibility claim and when it can't: A thought experiment can justify a possibility claim iff its evidence can be traced back to our knowledge of the actual world. Those that can are thereby justifiable by a low-flying modal thought experiment; those that cannot are unjustifiable. But as we've seen, there is reason to think that a significant

155

number of philosophically interesting philosophical claims can be evaluated by means of low-flying modal thought experiments.

However, this reply only partially mitigates the force of BAAAP. For while there may be reason to think that many philosophically interesting modal claims can be justified by low-flying modal thought experiments, there seem to be many that cannot. Happily, our framework provides the means for supplementing the above reply with a criticism of (2). For our framework puts us in a position to see that the apparent plausibility of (2) depends on a failure to appreciate that not all thought experiments depend for their utility upon whether they can justify a modal claim. More precisely, the apparent plausibility of (2) is based on two faulty assumptions: a faulty assumption about the sorts of facts thought experiments can reveal, and a faulty assumption about the sorts of claims thought experiments can properly evaluate. The former assumption is based on a failure to appreciate that in addition to modal thought experiments, there are also non-modal thought experiments, and the direct aim of each is different. The direct aim of a modal thought experiment is to determine whether two or more properties are companions, while the direct aim of a non-modal thought experiment is to determine whether two or more properties are constituents.

The second assumption is based on a failure to appreciate that not all philosophical claims are reducible to claims about companionship. For while strict conditionals and bi-conditionals are companion claims, Roy-conditionals and Fine-analyses are not; they are constituent claims. And while modal thought experiments are properly used to directly evaluate the companion claims of strict conditionals and bi-

conditionals, non-modal thought experiments are properly used to directly evaluate the constituent claims of Roy-conditionals and Fine-analyses. But to push our reply here further, we can say that high-flying non-modal thought experiments can play a role in defeating strict conditionals *indirectly*, depending on whether one can appeal to NEA to reason from "not constituents" to "not companions" in a given case, as we've seen in our evaluation of the two cases discussed in the second half of this chapter.

If the points sketched above are on track, then our framework provides a principled way to respond to BAAAP, and in a way that puts the classical Armchair Philosopher in a position to have her cake and eat it, too. For it enables her to resist BAAAP, but in a way that allows her to grant the full force of the case for Remoteness Modal Skepticism, and yet without it severely affecting the legitimacy of her practice.

157

BIBLIOGRAPHY

Balog, Katalin. "Conceivability, Possibility, and the Mind-Body Problem", *Philosophical Review* 108 (1999), pp. 497-528.

Bealer, George. "A Priori Knowledge and the Scope of Philosophy", *Philosophical Studies*, 81 (1996), pp. 121-142.

-"A Theory of the A priori", *Philosophical Perspectives*, 13 (1999), pp. 29-55.

-"Intuition and the Autonomy of Philosophy", in DePaul, Michael R. and Ramsey, William (eds.), *Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Inquiry* (Lanham, MD: Rowman and Littlefield, 1999), pp. 201-239.

-"Modal Epistemology and the Rationalist Renaissance", in Gendler, Tamar and John Hawthorne, eds., *Conceivability and Possibility* (Oxford: Oxford University Press, 2002), pp. 71-125.

Boghossian, Paul and Christopher Peackocke, eds. *New Essays on the A priori* (Oxford: Clarendon Press, 2000).

Bonjour, Laurence. *In Defense of Pure Reason* (Cambridge: Cambridge University Press, 1998).

-"What is it Like to Be Me (Instead of a Bat)?" Unpublished paper, http://faculty.washington.edu/bonjour/Unpublished%20articles/MARTIAN.html

Block, Ned and David Stalnaker. "Conceptual Analysis, Dualism, and the Explanatory Gap", *Philosophical Review* 108 (1999), 1-46.

Byrne, Alex. "Possibility and Imagination", *Philosophical Perspectives* 21 (2007), pp. 125-144.

Carnap, Rudolph. "Empiricism, Semantics, and Ontology", in Kim, Jaegwon and Sosa, Kim, *Metaphysics: An Anthology* (Malden, MA: Blackwell, 2000), pp. 13-22.

Casullo, Albert, ed. *A priori Justification* (New York: Oxford University Press, 2003).

-"Reid and Mill on Hume's Maxim of Conceivability", *Analysis* 39 (1979), pp. 212-219.

Chalmers, David. *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press, 1997).

-"Précis of *The Conscious Mind*", *Philosophy and Phenomenological Research*, Vol. LIX, No. 2, June 1999.

-"Materialism and the Metaphysics of Modality", *Philosophy and Phenomenological Research*, Vol. LIX, No. 2, June 1999, pp. 473-493.

-"Conceptual Analysis and Reductive Explanation", *The Philosophical Review* 110 (July, 2001), pp. 315-360.

-"Does Conceivability Entail Possibility?", in ed. T. Gendler and J. Hawthorne, *Conceivability and Possibility* (Oxford: Oxford University Press, 2002), pp. 145-200.

-"Epistemic Two-Dimensional Semantics", *Philosophical Studies* 118 (Mar., 2004), pp. 153-226.

-"The Components of Content". Home Page. <http://consc.net/papers/content.pdf>.

-"The Nature of Epistemic Space". Home Page. <http://consc.net/papers/espace.html>.

-"The Foundations of Two-Dimensional Semantics". Home Page. <http://consc.net/papers/foundations.html>.

-"On Sense and Intension". Home Page. <http://consc.net/papers/intension.pdf>.

Connee, Earl and Richard Feldman. *Evidentialism: Essays in Epistemology* (New York: Oxford University Press, 2004).

Connell, Richard. *Substance and Modern Science* (Houston, TX: Center for Thomistic Studies, University of St. Thomas, 1988).

-Davies, Martin and Brian Humberstone. "Two Notions of Necessity", *Philosophical Studies* 38 (1980), pp.1-30.

Della Rocca, Michael. "Essentialists and Essentialism", *The Journal of Philosophy* 93 (April, 2006), pp. 186-202.

DePaul, Michael and William Ramsey, eds. *Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Inquiry* (Lanham, MD: Rowman and Littlefield, 1999).

DeRose, Keith. "Epistemic Possibilities", *Philosophical Review* 100 (1991), 581-605.

Descartes, Rene. *The Philosophical Writings of Descartes*, 3 vols., translated by Cottingham, John, Robert Stoothoff, and Dugald Murdoch, vol. 3 including Anthony Kenny (Cambridge: Cambridge University Press, 1988).

Fine, Kit. "Essence and Modality", *Philosophical Perspectives vol. 8, Logic and Language* (1994), pp. 1-16.

Fiocco, M. Oreste. "Conceivability, Imagination, and Modal Knowledge", *Philosophy and Phenomenological Research* Vol. LXXIV No. 2 (March, 2007), pp. 364-380.

-"Conceivability and Epistemic Possibility", *Erkenntnis* 67 (2007), pp. 387-399.

Fischer, John Martin. "Responsibility and Alternative Possibilities", in his *My Way: Essays on Moral Responsibility* (Oxford: Oxford University Press, 2006).

Frankfurt, Harry. "Alternative Possibilities and Moral Responsibility", *Journal of Philosophy* 66 (1969), pp. 829-839.

Fumerton Richard. "Render Unto Philosophy That Which Is Philosophy's", *Midwest Studies in Philosophy*, XXXI (2007), pp. 56-67.

-"Personal Identity and Thought-Experiments", *The Philosophical Quarterly*, Vol. 52, No. 206 (2002), pp. 34-54.

-and John Hawthorne, eds. *Conceivability and Possibility* (Oxford: Oxford University Press, 2002).

Gertler, Brie. "A Defense of the Knowledge Argument", *Philosophical Studies* 93 (1999), pp. 317-36.

-"The Explanatory Gap is Not an Illusion: Reply to Michael Tye", *Mind* 110 (2001), pp. 689-694.

-"Conceptual Analysis, Explanatory Reduction, and Conceivability Arguments about the Mind", *Nous* 36 (March, 2002), pp. 22-49.

-"How to Draw Ontological Conclusions from Introspective Data", in Gertler (ed.) *Privileged Access: Philosophical Accounts of Self-Knowledge* (Ashgate, 2003).

Gettier, Edmund. "Is Justified True Belief Knowledge?", *Analysis* 23 (1963), pp. 121-123.

Hawthorne, John. "Advice for Physicalists". *Philosophical Studies* (May, 2002), pp. 17-52.

-"Intrinsic Properties and Natural Relations", *Philosophy and Phenomenological Research*, Vol. 63, No. 2 (Sep., 2001), pp. 399-403.

Hill, Christopher. "Imaginability, Conceivability, Possibility and the Mind-Body Problem", *Philosophical Studies* 87 (1997), pp. 61-85.

-"Modality, Modal Epistemology, and the Metaphysics of Consciousness", in Nichols, Shaun. *The Architecture of the Imagination* (New York: Oxford University Press, 2006), pp. 205-235.

Hoffman, Joshua and Rosenkrantz, Gary S. *Substance: Its Nature and Existence* (New York: Routledge, 1997).

Hoffman, Paul. "The Unity of Descartes's Man", *Philosophical Review* 95 (1986), pp. 339-370.

-"Cartesian Composites", *Journal of the History of Philosophy* 37 (1999), pp. 251-270.

-"Descartes's Theory of Distinction", *Philosophy and Phenomenological Research* (2002), pp. 57-78.

Hume, David. *A Treatise of Human Nature*, Selby-Bigge, L.A. (ed.), second edition edited by Nidditch, Peter H. (Oxford: Oxford University Press, 1978).

-*An Enquiry Concerning Human Understanding*, edited by Steinberg, Eric, second edition (Indianapolis, IN: Hackett, 1993).

Ichikawa, Jonathan and Jarvis, Benjamin. "Thought Experiment Intuitions and Truth in Fiction", *Philosophical Studies*, forthcoming.

Jackson, Frank. "Epiphenomenal Qualia", *Philosophical Quarterly* 32 (1982), pp. 127-136.

-"What Mary Didn't Know", *Journal of Philosophy* 83 (1986), pp. 291-295.

-"Postscript on Qualia", in Jackson, *Mind, Method, and Conditionals* (London: Routledge, 1998), pp. 76-79.

-"Metaphysics by Possible Cases", in Jackson, *Mind, Method, and Conditionals* (London: Routledge, 1998), pp. 133-153.

-"Armchair Metaphysics", in Jackson, *Mind, Method, and Conditionals* (London: Routledge, 1998), pp. 154-176.

-*From Metaphysics to Ethics: A Defense of Conceptual Analysis* (Oxford: Oxford University Press, 2000).

Jenkins, Carrie. "Modal Knowledge, Counterfactual Knowledge, and the Role of Experience", *The Philosophical Quarterly*, forthcoming.

Kind, Amy. "Putting the Image Back in Imagination", *Philosophy and Phenomenological Research*, Vol, 62, No.1 (Jan, 2001), pp. 85-109.

Kim, Jaegwon. "What is "Naturalized Epistemology?", in Sosa, Ernest and Kim, Jaegwon, *Epistemology: An Anthology* (Malden, MA: Blackwell, 2000), pp. 301-313.

Kripke, Saul. *Naming and Necessity* (Cambridge, Mass: Harvard University Press, 1980).

Kung, Peter. *Imagination and Modal Epistemology* Ph.D. dissertation, New York University, 2004.

-"Imaginability as a Guide to Possibility", ms.

Law, Stephen. "Loar's Defense of Physicalism", *Ratio* (new series) 17 (March, 2004), pp. 60-67.

Leon, Felipe and Tognazzini, Neal. "Why Frankfurt-Examples Don't Need to Succeed to Succeed", *Philosophy and Phenomenological Research* (forthcoming).

Lehrer, Keith. "Knowledge, Truth, and Evidence", *Analysis* 25:5 (1965), pp. 168-175.

Levine, Joseph. "Conceivability and the Metaphysics of Mind", *Nous* 32 (1998), 449-480.

-"Materialism and Qualia: The Explanatory Gap", *Pacific Philosophical Quarterly* 64 (1983), pp. 354-361.

Lewis, David. *On the Plurality of Worlds* (Oxford: Blackwell, 1984).

Loar, Brian. "Phenomenal States", *Philosophical Perspectives*, Vol. 4 (1990), pp. 81-108.

-"Phenomenal States (Second Version)", in Block, Ned, Flanagan, Owen J., and Guzeldere, Goven (eds.), *The Nature of Consciousness* (Cambridge, MA: MIT Press, 1997), pp. 597-616.

Lycan, William. "Vs. A New A priorist Argument for Dualism", *Philosophical Issues* 13, 2003, pp. 130-47.

-"Serious Metaphysics: Frank Jackson's Defense of Conceptual Analysis". Home Page. < http://www.unc.edu/%7Eujanel/Jackson2.htm>.

McGeer, Victoria. "The Trouble With Mary", *Pacific Philosophical Quarterly* 84 (2003), pp. 384-393.

Merricks, Trenton. "A New Objection to A priori Arguments for Dualism", *American Philosophical Quarterly* 31 (1994), pp. 81-85.

Nagel, Thomas. "What is it Like to Be a Bat?" *Philosophical Review* 83 (1974), pp. 435-450.

-Conceiving the Impossible and the Mind-Body Problem, *Philosophy*, Vol. 73, No. 285 (Jul., 1998), pp. 337-352.

Nichols, Shaun. *The Architecture of the Imagination: New Essays on Pretense, Possibility, and Fiction* (New York: Oxford University Press, 2006).

Nolan, Lawrence. "Descartes' Theory of Universals", *Philosophical Studies* 89 (1998), pp. 161-80.

Perry, John. *Knowledge, Possibility and Consciousness.* (Cambridge, MA: MIT Press, 2003).

Plantinga, Alvin. *The Nature of Necessity* (Oxford: Oxford University Press, 1974).

Putnam, Hilary. "The Meaning of 'Meaning'", in *Mind, Language and Reality: Philosophical Papers, Vol. 2* (Cambridge: Cambridge University Press, 1975), pp. 215-271.

Quine, Willard Van Orman. "Two Dogmas of Empiricism", in Poman, Louis (ed.), *The Theory of Knowledge: Classical and Contemporary Readings* (Belmont, CA: Wadsworth, 1993), pp. 395-407.

-"Epistemology Naturalized", in Sosa, Ernest and Kim, Jaegwon (eds.), *Epistemology: An Anthology* (Malden, MA: Blackwell, 2000), pp. 292-300.

Roy, Tony. "Worlds and Modality", *The Philosophical Review* 102 (1993), pp. 335-361.

-"Things and *De Re* Modality", *Nous* 34 (2000), pp. 56-84.

-"Making Sense of Relevant Semantics", unpublished ms., Nov. 16th, 2008. Department home page. <http://philosophy.csusb.edu/~troy/msrs-paper.pdf>

Schroeter, Laura. "The Limits of Conceptual Analysis", *Pacific Philosophical Quarterly* 85 (2004), pp. 425-453.

Seddon, John. "Logical Possibility", *Mind* 81 (1972), pp. 481-94.

Shoemaker, Sydney and Swinburne, Richard. *Personal Identity* (Oxford: Basil Blackwell, 1984).

Sidelle, Alan. *Necessity, Essence, and Individuation: A Defense of Conventionalism* (Ithaca, NY: Cornell University Press, 1989).

Smart, J.J.C. "Sensations and Brain Processes", *The Philosophical Review*, Vol. 68, N0. 2 (Apr., 1959), pp. 141-56.

Soames, Scott. *Reference and Description: The Case Against Two-Dimensionalism* (Princeton, NJ: Princeton University Press, 2005).

Sosa, Ernest. "Intuitions: Their Nature and Epistemic Efficacy", *Grazer Philosophische Studien* 74 (2007), pp. 51-67.

-"Experimental Philosophy and Philosophical Intuition", *Philosophical Studies* 132 (Jan., 2007), pp. 99-107.

Sosa, Ernest. "A Defense of the Use of Intuitions in Philosophy". Web Page. <http://homepage.mac.com/ernestsosa/FileSharing3.html>.

Stalnaker, Robert. "Varieties of Supervenience", *Philosophical Perspectives* 10 (1996), pp. 221-41.

-*Ways a World Might Be: Metaphysical and Anti-Metaphysical Essays* (Oxford: Oxford University Press, 2003).

Taliaferro, Charles. "Sensibility and Possibilia: A Defense of Thought Experiments", *Philosophia Christi*, Series 2, 3:2 (2001), pp. 403-420.

Thomas, Bruce. "Abstraction and the Real Distinction Between Mind and Body", *Canadian Journal of the History of Philosophy* (March, 1995), pp. 83-102.

Tidman, Paul. "Conceivability as a Test for Possibility", *American Philosophical Quarterly* 31 (1994), pp. 297-309.

-"The Justification of *A priori* Intuitions", *Philosophy and Phenomenological Research* 56:1 (Mar., 1996), pp. 161-171.

Trogdon, Kelly. "Monism and Intrinsicality", *The Australian Journal of Philosophy* (forthcoming).

Tye, Michael. "Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion", *Mind* 108 (1999), pp. 705-726.

-"Oh Yes it Is" *Mind* 110 (2001), pp. 695-697.

Vahid, Hamid. "Conceivability and Possibility: Chalmers on Modal Epistemology", *Philosophical Explorations* 9:3 (2006), pp. 243-260.

Vaidya, Anand. "Modal Rationalism and Modal Monism", *Erkenntnis* 68 (2008), pp. 191-212.

Van Cleve, James. "Conceivability and the Cartesian Argument for Dualism", *Pacific Philosophical Quarterly* 64, (1983), pp. 35-45.

Van Inwagen, Peter. "Ontological Arguments", *Nous* 11 (1977), pp. 375-95. Reprinted in *God, Knowledge, and Mystery: Essays in Philosophical Theology* (Ithaca and London), pp. 22-41.

-"Review of Richard Swinburne's *The Coherence of Theism*", *The Philosophical Review* 88 (1979), pp. 668-72.

-"Modal Epistemology", *Philosophical Studies* 92 (1998), pp. 67-84.

-"The Problem of Evil, the Problem of Air, and the Problem of Silence", *Philosophical Perspectives, Vol. 5: Philosophy of Religion* (1991), pp. 135-165. Reprinted in *God, Knowledge, and Mystery: Essays in Philosophical Theology* (Cornell UP, 1995), pp. 66-95.

-"Materialism and the Psychological Continuity Account of Personal Identity", *Philosophical Perspectives 11: Mind, Causation, and World* (1997), pp. 309-17. Reprinted in *Ontology, Identity, and Modality: Essays in Metaphysics (Cambridge UP, 2001)*, ch. 9.

-*Metaphysics*, 2nd. Ed. (Boulder, CO: Westview Press, 2002).
Weatherson, Brian. "Intrinsic Properties and Combinatorial Principles", *Philosophy and Phenomenological Research* Vol. 63, No. 2 (Sep., 2001), pp. 365-380.

-"What Good are Counterexamples?", *Philosophical Studies* 115 (2003), pp. 1-31.

Williamson, Timothy. *The Philosophy of Philosophy* (Oxford: OUP, 2007).

Witmer, Gene D., Butchard, William, and Trogdon, Kelly. "Intrinsicality Without Naturalness", *Philosophy and Phenomenological Research* Vol. LXX, No. 2 (March 2005), pp. 326-350.

Woolhouse, R.S. "From Conceivability to Possibility", *Ratio*, 14 (1972), pp. 144-54.

Worley, Sara. "Conceivability, Possibility and Physicalism", *Analysis* 63 (2003), pp. 15-23.

Yablo, Stephen. "Is Conceivability a Guide to Possibility?", *Philosophy and Phenomenological Research* 53 (1993), pp. 1-42.

-"The Real Distinction Between Mind and Body", *Canadian Journal of Philosophy*, Supplementary Volume 16, pp. 149-201.

-"Identity, Essence, and Indiscernability", *The Journal of Philosophy* 84 (1987), pp. 293-314.

-"Textbook Kripkeanism and the Open Texture of Arguments", *Pacific Philosophical Quarterly* 81 (2000), pp. 98-122.

-"Concepts and Consciousness", *Philosophy and Phenomenological Research* 59 (1999), pp. 455-463.

-"Coulda, Woulda, Shoulda", in Gendler Szabo, Tamar and John Hawthorne, *Conceivability and Possibility* (Oxford: OUP, 2001), pp. 441-491.

Zimmerman, Dean. "Two Cartesian Arguments for the Simplicity of the Soul". *American Philosophical Quarterly* 28 (1991), pp. 217-226.

Appendix I:
The Failure of Cartesian Modal Epistemology

In many places, Descartes justifies modal claims by appealing to his ability to conceive of things as being such-and-such. In doing so, he depends on a robust connection between conceivability and metaphysical possibility: If something is conceivable (in a certain sense, to be spelled out below), then it is metaphysically possible. Call this the 'conceivability-possibility connection' ('CPC' for short). But is such a connection really strong enough to warrant his judgments about the modal properties of things? One may have doubts today. A number of contemporary philosophers have advanced arguments that seem to weaken CPC.[91] For example, Peter Van Inwagen argues that just about any philosophically interesting modal claim is such that it cannot be justified by means of *a priori* thought experiments. But if this is the case, then it seems that CPC is undercut. How does Descartes' account of CPC hold up in light of contemporary criticisms? It is my contention that such modal claims retain significant positive epistemic status. However, I don't pretend to remove all of the problems for Descartes' CPC. Rather, I have the more modest aim of leaving it in a little better shape than when I found it.

This paper has seven sections. In the first section, I briefly lay out the relevant components of Descartes' basic metaphysics of substance, as well as his theory of

---

[91] See, e.g., Van Inwagen, Peter. "Modal Epistemology", *Philosophical Studies* 92 (1998), pp. 67-74.

distinction. In section II, I take a first pass at a discussion of the justification of the

inference from conceivability to possibility. In section III, I discuss a number of thorny

problems for justifying this inference. From that point, I spend the next three sections

(IV-VI) showing that Descartes has the materials to overcome these problems. Finally, in

section VII, I critically evaluate the defense of CPC. I conclude that his account of modal

epistemology is crucially dependent upon his account of corporeal substances, and since

the latter is now known to be false, his epistemology of modality is likewise seen to be

inadequate.

1. The Basic Picture: Metaphysics and the Theory of Distinction

Before we discuss Descartes' account of the ways in which things may be distinct, we

must first lay out several of his key metaphysical concepts: his notions of *substance*,

attribute, and *mode*. By *substance*, Descartes means any entity that can exist apart from

any other thing.[92] Descartes has two broad notions of substance, corresponding to two

notions of separate existence. Something is a finite substance iff it can exist apart from

any other entity besides God. And something is an *infinite* substance iff it can exist apart

from any other entity whatsoever.[93] Paul Hoffman has argued that at least one of

Descartes' notions of substance is extremely weak: A thing's qualifying as a substance

need merely imply that it can exist without existing as in a subject, unlike modes and

---

[92] *The Philosophical Writings of Descartes, vol. I*, trans. John Cottingham, Robert Stoothoff and Dugald Murdoch (Cambridge: Cambridge UP, 1985), p. 210 (henceforth abbreviated as 'CSM I').

[93] CSM I, 210. I was helped here, as well as with respect to the following discussion of Descartes's ontology, by Lawrence Nolan's paper, "Descartes' Theory of Universals", *Philosophical Studies* 89 (1998), pp. 161-80.

attributes.[94] Hoffman has also shown that Descartes' notion of separate existence is ambiguous, as we will see. We will have more to say about this when we discuss features of Descartes' epistemology, but for now, what I have said is sufficient.

By *attribute*, Descartes means a feature of a substance that is unchanging.[95] For example, any material object has the attribute of shape.[96] And no mind can exist without having the capacity for thought. Thus, if a substance lost an attribute, that substance would cease to exist. Now every substance has what Descartes calls its 'principal attribute'.[97] Such an attribute is a substance's most fundamental feature *qua* the kind of substance it is. There are two types of principal attribute: thought and extension. These correspond to and determine the two basic kinds of substance: mind and body. Thus, the principal attribute of any given mind is thought, and the principal attribute of any given body is extension.[98] All the other variable (or changeable) features of a given substance presuppose its principal attribute.[99]

---

[94] See, for example, Hoffman, Paul. "The Unity of Descartes's Man", *The Philosophical Review*, No. 3 (July 1986), p. 339.

[95] Cf. CSM I, 211-212.

[96] This example is based on that of Lawrence Nolan's paper, "Descartes' Theory of Universals", *Philosophical Studies 89* (1998), p. 163: "Descartes' famous piece of wax from the Second meditation provides a rich source of examples...Shape as a determinable property is an attribute of the wax, but the determinate shape that it has at any particular time is a mode."

[97] CSM I, 210-211.

[98] CSM I, 210.

[99] *Ibid.* The preceding discussion of substances and attributes leaves out an important topic: *composite* substances. Paul Hoffman has many helpful and illuminating things to

By *mode*, Descartes means any changeable feature of a substance. For example, my dog is an extended substance, and the particular size of my dog is one of its modes. That substance (my dog) would (or could) continue to exist if it grew several inches in height.

We must also discuss Descartes' epistemological notions of *exclusion* and *abstraction*.[100] By *abstraction*, Descartes means an operation of the intellect whereby one focuses one's attention on a particular feature of a thing, though the other features need not be totally ignored. So, for example, I may attend to my thought of a book, focusing my attention on its rectangular shape. In doing so, I am abstracting it away from its other features, such as its thickness, its size, etc. By way of contrast, *exclusion* is an operation of the intellect, whereby one attends to a pair of features of an idea of a particular thing – say, F and G – and coherently conceives of that thing as possessing F while simultaneously not possessing G. So, for example, consider my idea of a particular book again. I may attend to its feature of being made of paper, while coherently conceiving of it as existing even though it is an inch narrower. In such a case I have *excluded* the particular size of the book from my conception of the book. As we will see, exclusion plays a crucial role in Descartes' theory of distinction.

---

say on this topic in (e.g.) his papers "Cartesian Composites", *Journal of the History of Philosophy* 37, No. 2 (April, 1999), pp. 251-270. and "The Unity of Descartes's Man". However, I believe we may safely leave this discussion aside for the purposes of this paper.

[100] The following discussion of these two notions is indebted to Bruce Thomas' discussion in his "Abstraction and the Real Distinction Between Mind and Body", *Canadian Journal of the History of Philosophy* (March, 1995), pp. 83-102.

According to Descartes, there are three ways in which things may be distinct:

really, modally, and rationally.[101] Each distinction can be can be defined in

epistemological terms and metaphysical terms.[102] That is, each distinction can be defined

in terms of the ways in which the *presentational contents of an idea* of a thing (or things)

can be excluded from one another (the epistemological versions of the definitions), or in

terms of the ways in which *the things themselves* can be separated from one another (the

metaphysical versions of the definitions). To keep them separate, I will add a subscripted

'E' to each expression denoting a distinction that is defined in epistemological terms, and

I will add a subscripted 'M' to each expression denoting a distinction defined in

metaphysical terms. Let's start with the former definitions. Two things are *really*

distinct$_E$ iff each can be clearly and distinctly conceived to exist apart from the other. [103]

Using the notion of exclusion mentioned above, we may express his notion of the real

distinction as follows: Consider any ideas I_1 and I_2, presentational contents a and b in I_1

and I_2, respectively, and putative entities A and B presentationally represented in I_1 and I_2

(again, respectively).[104] Then we may say that A and B are really distinct$_E$ iff one can

exclude a from b in I_1 and I_2 and vice-versa.

---

[101] CSM I, 213-214. I was greatly helped in this section by Paul Hoffman's "Descartes' Theory of Distinction", *Philosophy and Phenomenological Research*, Vol. LXIV, No. 1, January 2002, pp. 57-78, esp. pp. 57-60.

[102] Cf. Hoffman, Paul. "Descartes's Theory of Distinction", pp. 57-60.

[103] CSM I, 213.

[104]On Descartes' account of exclusion, we aren't necessarily excluding one presentational content, a, in an idea from another content, b, *in the same idea*. Rather, Descartes seems to think that exclusion can take place among the contents of more than one idea, and

As I mentioned above, Descartes' notion of separate existence is ambiguous. Paul Hoffman points out that, at least in the context of the real distinction, the heart of Descartes' notion of separate (or separable) existence is that of the ability of one thing to exist *qua* the kind of thing it is without the principal attribute of another thing existing in it.[105] This construal of the locution is preferable to others for a number of reasons, but here it is enough to mention one: such a notion doesn't imply the controversial idea that really distinct things can exist "all by themselves", as it were. This is the consequence that we want, since it's not clear that the mutual excludability of our ideas of things implies the (possible) existential independence of the things they represent. Thus, let's follow Hoffman in adopting this construal of 'can exist apart'. With that said, let's move on to Descartes' notion of a modal distinction.

There are two ways in which things may be modally distinct. Call them 'modally$_1$ distinct$_E$' and 'modally$_2$ distinct$_E$'. Two things, A and B, are *modally$_1$ distinct$_E$* iff one's corresponding ideas of A and B, a and b, are such that one can exclude a from b, but one cannot exclude b from a, or vice-versa (where 'or' is taken in the exclusive sense).[106] Two things, A and B, are *modally$_2$ distinct$_E$* iff a can be excluded from b and b can be excluded from a, though neither can be excluded from the idea of a third thing c

these ideas need not be encompassed in some larger idea. For textual support, see Descartes' Letter to Gibieuf, 19 Jan., 1642, in *The Philosophical Writings of Descartes, vol. III*, trans. John Cottingham, Robert Stoothoff, Dugald Murdoch and Anthony Kenny (Cambridge: Cambridge UP, 1991), pp. 201-202. I owe this point, and the reference, to Paul Hoffman.

[105] Hoffman, Paul. "Descartes's Theory of Distinction", p. 69.

[106] Cf. CSM I, 213-214.

(though c can be excluded from both a and b).[107] Finally, two things, A and B, are *rationally distinct$_E$* iff it is not the case that either a can be excluded from b or b can be excluded from a.[108]

Corresponding to the above epistemological definitions are several metaphysical definitions. We may say that any two things, A and B, are really distinct$_M$ iff A can exist without B and B can exist without A. And A and B are modally$_1$ distinct$_M$ iff either A can exist without B and B cannot exist without A; or B can exist without A and A cannot exist without B (but not both). Furthermore, A and B are modally$_2$ distinct$_M$ iff A can exist without B and B can exist without A, but neither A nor B can exist without some further thing, C (though C can exist without either A or B). Finally, A and B are rationally distinct$_M$ iff neither A nor B can exist without the other.[109]

With the above sketch of Descartes' key metaphysical and epistemological theses, we are now in a position to state his theory of distinction (TOD):

> (TOD): For any ideas I_1 and I_2, presentational contents a and b in I_1 and I_2, respectively, and putative entities A and B presentationally represented in I_1 and I_2 by a and b (again, respectively), if a and b are clearly and distinctly perceived, then if the a and b differ in presentational content, then A and B are either really distinct$_M$, modally distinct$_M$ or rationally distinct$_M$.

At this point, one may wonder what lies within the scope of clear and distinct perception. Paul Hoffman has argued persuasively that its scope is (to paraphrase him radically) the set of all ideas and presentational contents that are of particulars (whether

---

[107] Cf. CSM I, 214.

[108] Cf. CSM I, 214-215. In this passage, the translators use the locution 'conceptual distinction' for what I am here calling a 'rational distinction'.

[109] Descartes discusses these theses in the passages mentioned above in CSM I, 213-214.

173

of objects or instances of features of things, such as attributes and modes), including what are traditionally called the "secondary qualities" (i.e., colors, sounds, etc.), so long as the latter aren't considered as entities existing in objects outside of us.[110] Let's follow Hoffman on this and limit the scope of clear and distinct perception accordingly.

2. From Conceivability to Possibility: A First Pass

Now that we have Descartes' version of TOD before us, the crucial issue of justification arises: what justifies Descartes' move from the epistemological theses to the metaphysical ones? Descartes gives us his rationale here with respect to the real distinction when he argues that the mind is really distinct from the body:

> God can bring about whatever we clearly and distinctly perceive in a way exactly corresponding to our perception of it...But we clearly and distinctly perceive the mind, that is, a thinking substance, apart from the body, that is, apart from an extended substance...And conversely we can clearly and distinctly perceive the body apart from the mind (as everyone admits). Therefore the mind can, at least through the power of God, exist without the body; and similarly the body can exist apart from the mind.[111]

From this passage, we may extract Descartes' general defense of TOD with respect to things that are really distinct:

1. Anything that one can clearly and distinctly conceive is such that God is able to bring it about in a way exactly corresponding to one's conception of it.
2. I can clearly and distinctly conceive of the existence of A apart from the existence of B, and vice-versa.
3. Therefore, God is able to bring it about that A exists apart from B and vice-versa.
4. Something is metaphysically possible iff it is such that God is able to bring it about.
5. Therefore, the existence of A apart from the existence of B, and vice-versa, is

---

[110] See his "Descartes's Theory of Distinction", esp. pp. 62-65.

[111] CSM II 119-120.

174

metaphysically possible.

6. A and B are really distinct$_M$ iff the existence of A apart from the existence of B, and vice-versa, is metaphysically possible.
7. Therefore, A and B are really distinct$_M$.

Premises (1) and (4) appear to be the core grounds for Descartes' inference from conceivability to possibility. Together, they entail a straightforward version of CPC:

8. (CPC) Anything that is clearly and distinctly perceivable is really (i.e., metaphysically) possible.

And (1), a key premise for the derivation of CPC, seems to imply that metaphysical possibility depends crucially on the omnipotence of God. However, this turns out to be merely apparent. For as Descartes tells us:

> I introduce the power of God as a means to separate mind and body not because any extraordinary power is needed to bring about such a separation but because the preceding arguments have dealt solely with God, and hence there was nothing else I could use to make the separation. Our knowledge that two things are really distinct is not affected by the nature of the power that separates them.[112]

This statement would seem to imply that the omnipotence of God is not crucial to the inference from conceivability to possibility. Rather, the crucially relevant factor is the logical or conceptual coherence of a given clear and distinct conception: All coherently conceived things (or states of affairs) are metaphysically possible. But if so, then if we

---

[112] *Ibid.*, p. 120. Descartes seems to be implying the same point when he says, "It remains for me to examine whether material things exist. And at least I know they are capable of existing, in so far as they are the subject-matter of pure mathematics, since I perceive them clearly and distinctly. For there is no doubt that God is capable of creating everything I am capable of perceiving in this manner; *and I have never judged that something could not be made by him except on the grounds that there would be a contradiction in my perceiving it distinctly.*" (CSM II 50. Italics mine.) For the portion italicized suggests that it is something about the represented features themselves and their relations to other features, that is the crucial issue with respect to modality.

175

modify (1) and (4), and add a few auxiliary premises, then we may derive (8) (i.e., CPC) as follows:

> 1'. For any x[113], and conception, C, of x, if C is clear and distinct, then (i) the presentational contents of C are all logically and/or conceptually consistent with one another, and (ii) x is intelligible[114] if conceived by means of C.

> M. For any x, if (i) the presentational contents of a conception of x are logically and/or conceptually consistent with one another, and (ii) x is intelligible (if conceived by means of an accurate conception of it), then the constituents of x are such that (a) they are compatible with one another and (b) they constitute a complete thing[115] if co-instantiated.

---

[113] Here and following, it is to be understood that the scope of the universal quantifier for the variable 'x' is to be wide enough so as to include not only (e.g.) substances, but also states of affairs.

[114] I will speak more about intelligibility in the next footnote, where I briefly discuss Descartes' notion of a complete thing.

[115] Roughly, x is a *complete thing* (or, x is a *substance*) iff x (a) x lacks nothing that would prevent x from existing without a subject (even by a miracle). Descartes appears to define the notion of a complete thing in *epistemological* terms; specifically, in terms of *intelligibility*: x is a *complete thing* iff [for any accurate conception, I, of x, and presentational content c, if I requires c for the intelligibility of x, then I has c]. Typically (and perhaps in every case), rendering a conception of a thing unintelligible amounts to excluding a presentational content that corresponds to an *attribute* of that thing. At least, this seems to be the case with respect to conceptions of *simple* substances. For example, if one excludes the concept of shape from one's conception of a piece of wax, one will render the conception of the wax unintelligible. The metaphysical implication is then supposed to be that the wax is not a complete thing without the attribute of shape, and so cannot exist without it (even by a miracle). However, one may exclude a presentational content corresponding to an attribute of a *composite* substance, and still be left with an intelligible conception of a thing (namely, the conception of the remaining, unexcluded substance(s)). Of course, in cases of the latter type, the conception of the *composite* substance would be rendered unintelligible after such an operation of the intellect. I believe this sketch of an account captures Descartes' thought on the matter when he says: "...by a 'complete thing' I simply mean a substance endowed with the forms or attributes which enable me to recognize that it is a substance." (CSM II, 156); and when he says, "...if we subsequently wanted to strip the substance of the attributes through which we know it, we would be destroying our entire knowledge of it." (*Ibid.*). Of course, (at least) this latter quote is referring to the *principal* attribute of a thing, and what would happen if we excluded our conception of it from our conception of a substance by an operation of

176

From (1') and (M) we get

N. For any x, and conception, C, of x, if C is clear and distinct, then the constituents of x are such that (a) they are compatible with one another, and (b) they constitute a complete thing if co-instantiated.

Finally, if we add

4'. For any x, x is (metaphysically) possible iff the constituents of x are such that (a) they are compatible with one another and (b) they constitute a complete thing if co-instantiated.

Then from (N) and (4') we get

8. (i.e., CPC) For any x, and conception, C, of x, if C is clear and distinct, then x is (metaphysically) possible

In this way, then, we may derive (8) without the need to make explicit reference to omnipotence in order to get from conceivability to possibility. Let me reiterate the fundamental justification for CPC, but in a slightly different way. Possibility is grounded in (a) the *compatibility* of the conceived constituents, and (b) the *completeness* of the constituents when taken together. But clear and distinct conceptions (the "conceivability" side of the equation) "mirror" their referents, as it were. They are perfect (or in any case sufficiently accurate) representations of the things they represent. To get this point across a bit more precisely, consider the following expression of the current account of conceivability, and of possibility, respectively:

---

the intellect. However, it seems clear that a similar point would apply if we excluded a conception of an attribute that is merely rationally distinct from a principal attribute. These (extremely rough) construals of Descartes' notions of a complete thing and of intelligibility are strongly influenced by the discussion of them in the following papers: Hoffman, Paul. "Cartesian Composites", *Journal of the History of Philosophy* 37, no. 2, pp. 251-70; Thomas, Bruce M. "Abstraction and the Real Distinction Between Mind and Body" *Canadian Journal of Philosophy* (March, 1995), pp. 83-101.

177

Conceivability: If something is *conceivable*, then (i) the conception contains each of the minimal constituent concepts required to make the conception of the thing intelligible, and (ii) these constituent concepts are logically and/or conceptually consistent with each other.

Possibility: If something is (metaphysically) *possible*, then (i') it possesses each of the minimal constituents required for it to be instantiated (i.e., it is complete), and (ii') these features are compatible with one another.

In light of these accounts, we may say that clear and distinct conceptions mirror objects and/or states of affairs in such a way that (i) is a sufficiently accurate mental analogue of (i') and (ii) is a sufficiently accurate mental analogue of (ii').

3. Problems for the Justification of CPC, and a Strategy for their Solution

Unfortunately, at this point it is evident that our account of the justification of the move from conceivability to possibility is inadequate with respect to our present interests. For our original question was in regard to the move from conceivability to possibility *in TOD*. And TOD is a principle that is supposed to be useful for analyzing the modal properties of *actual* objects. However, while the satisfaction of the premises above may be sufficient to justify claims about what's possible with respect to (currently) *non-actual* objects and states of affairs (should they be actualized), they don't seem to be sufficient to justify claims about the modal properties of (currently) *actual* objects. For whatever one wants to say about the former types of cases, there is no doubt that, when it comes to the latter types of cases, there can be a *discrepancy* between the way we conceive of an object (or state of affairs) and the way the object actually is. But if so, then the following scenario might take place: We form a conception of an actual object, say, a piece of wax. We then mentally "strip away" what we *take to be* only its accidental features, with the

178

result that our conception of the wax (a) is complete and (b) its features are all

compatible. However, unbeknownst to us, we have stripped away an essential feature of

the wax, such that, while our new conception may denote a genuinely possible (though

probably non-actual) object, the object represented is no longer *the actual wax*[116]. And as

a result of this, we make a faulty inference with respect to what is a possible state of

affairs for *the actual wax*.[117] For the latter types of cases, then, we need a modified

version of CPC that ensures that the conception of the entity denoted by the 'x'

accurately correspond to *the actual, existent entity* conceived. Therefore, we will have

two versions of CPC: one that applies to non-actual, (perhaps merely) conceptually

constructed entities and one that applies to actual objects. The latter can be expressed as

follows:

> (CPC*) For any **actual** x, and conception, C, if C is a clear and distinct
> conception **of x** (as opposed to (say) some intelligible but non-actual entity that
> resembles x), then the existence of x (in the state in which it is conceived) is
> metaphysically possible.[118]

Unfortunately, with this new criterion, it seems that we have lost any precise sense of

what a "clear and distinct conception of x" amounts to. For we have just been forced to

---

[116] Descartes would likely deny that the "stripping away" of an essential feature of an
object could generate a conception of a genuinely possible object. However, as I will
argue later, such a denial depends upon a certain account of the nature of simple extended
objects, and this account is (at least currently) extremely controversial at best. In any
case, consider the above worry as one raised by a person who denies Descartes' account
of the nature of simple extended objects.

[117] This also shows that, in order to avoid counterexamples, we must restrict the scope of
the quantifiers that range over the xs in the *earlier* premises (especially CPC) to *non-
actual* entities and states of affairs.

[118] To prevent making this paper intolerably long, I will leave out analogues to premises
(1'), (4'), and auxiliary premises M and N, from which way may infer CPC*.

179

concede that it goes beyond its having compatible constituents that, when taken together, constitute a complete thing. What else is needed? The obvious first response is that the conception must also accurately correspond to the actual thing conceived. But this is of little help; for we don't yet know what conditions have to be met for this to be true of a given conception. So we need a criterion for determining when a conception meets this and the other two conditions of accuracy; something of the following form (call any criterion that fleshes out the following form 'CAS' (for "the Criterion of (CPC*) Antecedent Satisfaction".) :

> (CAS) For any **actual** x, and conception, C, (putatively) of x, C is a clear and
>
> distinct conception **of x** iff _____.

It is at this point that things get difficult. For suppose we grant that CPC* is true, and that its truth grounds the move from conceivability to possibility in TOD. Then we are immediately confronted with two problems: (1) we haven't been given a criterion for determining **when** the antecedent of this conditional has been satisfied. But if not, then it appears that any modal claim based upon an application of CPC* will be unjustified. Furthermore, (2) we have no antecedent reason to think **that** the antecedent of the conditional *can* be satisfied: perhaps none of our conceptions accurately represent the natures of actual objects. But if the antecedent of CPC* can't be satisfied (or at least, if we can't be confident that it is satisfied), then (at least, for all we know) CPC* is trivially true, in which case, again, none of our modal claims are justified. It will prove extremely helpful in answering these two objections if we look at them fairly closely.

180

To motivate and clarify problem (1) a bit more, I will discuss two kinds of cases that seem to be problematic for CPC* as it stands. Before we begin, however, I need to lay out an artificial vocabulary. Let: 'O' denote an arbitrary object or state of affairs; 'A', 'B', 'C', 'D' denote the possible constituent essential features of an object; 'I' and 'J' denote arbitrary conceptions of things; 'a', 'b', 'c', 'd' denote the possible constituent concepts of a conception, such that a represents A, b represents B, etc.; and let '↔' denote the material biconditional.

> *"Deep Structure" cases*: Cases where, although *the features that one grasps* are primary and essential to the thing conceived, and one can be sure that the features grasped *are* essential, at least one of the essential features of the thing is beyond one's grasping (henceforth, let's say that such features are *cognitively hidden*):

Suppose that an agent has the following **apparent** conception of O:

$$I: \quad a \leftrightarrow b \leftrightarrow c$$

(I will express the importance of the biconditional shortly) Suppose further that a **true** conception of O diverges from the apparent conception I, since O has the following essential features:

$$O: \quad A \quad B \quad C \quad D$$

Thus, while I is accurate *as far as it goes*, it isn't *completely* accurate. For it lacks the constituent concept that would make it perfectly accurate, viz., d (which represents D of O). Thus, the true conception of O can be expressed as follows:

$$J: \quad a \leftrightarrow b \leftrightarrow c \leftrightarrow d$$

Finally, suppose that we humans are incapable of attaining to the true conception of O. In such a case, if, based on the *apparent* analysis, one tries to make inferences about what

181

is possible with respect to O, then one may well go wrong in such an inference. For example, based on the *apparent* conception of O – i.e., I - one may infer that since I does not presuppose d, nor does the latter presuppose the former, that therefore O can exist apart from D. However, since (by hypothesis) the *true* conception is J, it follows that the proposed modal statement is false.

It is important to note the following feature of Deep Structure Cases. Notice that the constituent concepts of the apparent conception (and the true conception) are related to each other by a biconditional. This guarantees that **if** the person entertaining the apparent concept can see that such a biconditional obtains between the constituent concepts, then he can be sure that *at least* the referents of *those* constituents are essential to the thing conceived. (Henceforth, let us say that if a biconditional obtains between any two concepts referring to a conceived thing, then those concepts are *logically connected.*) Unfortunately, however, in Deep Structure cases, the person entertaining the concept cannot see at least one other constituent concept of an essential feature of the thing, even though it is logically connected to the others. We may therefore characterize Deep Structure cases as those in which the constituents of a conception of a thing are all logically connected, but at least one essential feature of the thing is cognitively hidden.

> *"Lawless Features" cases*: Cases where (i) one has a conception C, such that C captures the essential features of a thing A, but (ii) at least one constituent concept of C is not logically and/or conceptually related to the other essential features of A (henceforth, call such constituent concepts 'logically unconnected').

Suppose that one has the true conception of O, and that the true conception of O has the following analysis

182

I: a ↔ b ↔ c   d

So I is a true conception of O in virtue of the fact that O has the following constituents for its essential features:

O: A   B   C   D

Now notice that d, a constituent concept of I, isn't logically connected to the other constituent concepts of I, though d fulfills its role by correctly representing D, an essential feature of O. In such a case, one is able to conceive of all of the essential features of a thing, yet one is unable to see at least one essential feature (in this case, D) as an essential feature of a thing. Because of this, one seems unable to guarantee that D is essential. But if not, then one is liable to make false inferences about what is possible with respect to O. For example, one may reason that, since the rest of the concept I isn't rendered unintelligible when one excludes d from it, one can justly infer that D is not an essential feature of O, and so O can exist without D. However, since (by hypothesis) this is false, it follows that the proposed modal statement is false. Therefore, we see that Lawless Features Cases cause trouble for CPC* in two ways. First, one is liable to go from a true conception of a thing to a false one. For although all of the essential features of the thing can be grasped by the mind, such features aren't logically and/or conceptually interconnected in a way that would make their intelligible exclusion impossible: they aren't logically connected in the way defined above. Compare this case with Deep Structure Cases. Recall that in such cases, the person is able to become certain that the constituents of her concept of a thing represent some of the essential features of that thing in virtue of the facts that (i) such constituents are logically connected to the

183

others and (ii) she can see that this is so. By way of contrast, in Lawless Features Cases, at least one constituent concept is not logically connected to the others. To the extent that this is true, the person cannot be sure that such a constituent is part of the true conception of the essence of the thing. Second, even if one sticks with the original, true conception, one can't be sure that the conception is true, and for the reasons just discussed. But if not, then even if one bases modal inferences upon such a concept, those inferences are unjustified. We may characterize Lawless feature cases as those in which all of the essential features of a thing are cognitively accessible, but at least one constituent concept of the conception of the thing is logically unconnected to the other constituents.

Call Lawless Features Cases and Deep Structure Cases 'the Problem Cases'. Then we may summarize the worries that arise for CPC* in light of the preceding discussion as follows. Since we don't yet have an account of (CAS), then if we try to use CPC* without it, then we can't yet rule out that any of our conceptions of things aren't instances of the Problem Cases. But if not, then we can't use it to make fully justified *de re* modal statements.[119]

In light of the preceding, it is clear that if Descartes is to be able to make fully justified inferences from conceivability to possibility, then problems (1) and (2) must be answered. Also in light of the preceding, we see that this amounts to doing the following two things. First, the Problem Cases - Deep Structure Cases and Lawless Features Cases (or hybrids of them) - must be prevented from counting as satisfying the antecedent of CPC*. I suggest that meeting this requirement can be accomplished if (a) we can come up

---

[119] Here and henceforth, let 'fully justified' (and similar locutions) denote the kind of indubitable certainty that Descartes was after.

with a *criterion* that prevents such cases from satisfying the antecedent of CPC*. This will amount to putting flesh on the bones of our schema for CAS. Second, CPC* must not be trivially true in the actual world. I suggest that if (b) Descartes had a reliable *method* for forming conceptions of things that satisfy CAS (and so, satisfy the antecedent of CPC*), and (c) Descartes has an *ontology*, such that we can form fully accurate conceptions of at least some of its members, then this requirement can be met. My overall strategy for answering problems (1) and (2), therefore, amounts to defending the following three claims:

> (i) There is a **criterion** that, when utilized, ensures that one's conception of a thing is complete and accurate.

> (ii) There is a reliable **method** to gain such conceptions

> (iii) There are **objects** in Descartes's ontology, such that we can form complete and accurate conceptions of their natures by means of this method.

The next three sections will be devoted to defending these three claims.

4. Defending (i): The Criterion

To see what criterion would do the job, we must first get a bit clearer on the specific kinds of conditions in which the Problem Cases arise. In Deep Structure Cases, the bad modal inference is made possible by the fact that a concept of an essential feature of a thing is *cognitively inaccessible* yet *logically connected*[120] to the true conception of that thing. Now the heart of the problem with such cases is this. If you don't have a complete

---

[120] We will have more to say about some important ways in which the constituents of the concept of a thing may be logically interconnected a bit later. But for now, let this promissory note suffice. If it does *not* suffice, then I petition the reader to skip ahead to the discussion of weak and strong analyticity on pp. 27-28 before reading any further.

185

grasp of the principal attribute of a thing, then you can't be sure of the exact conditions in

which a thing instantiating that nature can exist. But if not, then if you make modal

claims about such a thing, you are liable to say something that is incompatible with those

conditions. Thus, to avoid such cases, Descartes must provide a criterion of justified

modal inference (from conceivability to possibility)[121] that has a scope limited to entities

that satisfy the following condition:

> *The Shallow Structure Condition (SSC)*: (i) All of the essential features of
> things (and, in general, any features relevant to assessing modal claims)
> are within our cognitive grasp, and (ii) there is a way to be sure whether or
> not (i) is satisfied.

In Lawless Features Cases, a bad modal inference is made possible by the fact

that a concept of an essential feature of a thing is *cognitively accessible* yet *logically*

unconnected (in the way mentioned above) to the other constituents of the true

conception of that thing. Now the heart of the problem of such cases is this. In order to

meet Cartesian standards of justification, we need to be certain that a feature of a thing is

essential to it. If the following two conditions are true with respect to an arbitrary thing

and its features, then Cartesian standards of justification can be met[122]: (i) its concept is

logically connected to the concepts of the other essential features of that thing, and (ii) we

---

[121] Henceforth, let's always assume that this parenthetical qualification applies to the
locution, "criterion of justified modal inference".

[122] While I wish to say that the satisfaction of these conditions is *sufficient* for meeting
Cartesian standards of justification, I do not wish to claim that satisfying these conditions
is also *necessary*. That is, I do not want to deny that there may be *other* ways in which
Cartesian standards of justification can be met. For example, Paul Hoffman has
suggested to me that perhaps absolute certainty could be achieved if the constituent
concepts are connected as those involved in Kantian synthetic *a priori* propositions.
Indeed, I allow that the logical connections mentioned in my two conditions may not all
be analytic.

186

see *that* it's logically connected to the other concepts of the essential features of that thing. Conversely, such certainty is impossible if there is no relevant logical connection between the concept of one essential feature and the others, or if there is but we can't see this is so. Therefore, to avoid Lawless Features cases, Descartes must provide a criterion of justified modal inference that has a scope limited to entities that satisfy both SSC and the following condition:

> *The Ideal Conditions Condition (ICC)*: (i) All of the constituents of the concept of a thing are logically connected, and (ii) we can tell when (i) is satisfied.

In light of the above discussion, we may now state succinctly what Descartes' Criterion of (CPC*) Antecedent Satisfaction (CAS) must do. First of all, it must not allow instances of Deep Structure Cases and Lawless Features Cases to count as conceptions that satisfy the antecedent of CPC*. We have seen that this amounts to providing a criterion that entails that a conception C satisfies the antecedent of CPC* only when C satisfies both SSC and ICC: C must be such that (a) it captures all of the essential features of its referent, i.e., they are all are *cognitively accessed* (and we can be sure that this is so), and (b), all of the constituents of C are *logically connected* (and we can tell when this is so). Thus, we may now state the criterion for satisfying the antecedent of CPC* (CAS), which will complete our case for 2(a):

> (CAS) For any **actual** x, and conception, C, (putatively) of x, C is a clear and distinct conception **of x** iff C satisfies SSC and ICC.

> (Corollary) A conception, C, of an actual entity, x, satisfies SSC and ICC iff for any essential feature, F, of x, (a) F is *cognitively accessed* in virtue of its being accurately represented in C (and we can be sure that this is so) and (b) for any

187

distinct constituents, a and b, of C, a is *logically connected* to b (and we can be sure that this is so).

5. Defending (ii): A Method of Grasping Essences

With our criterion in hand, it now needs to be shown that Descartes has a method for conceiving of actual objects or states of affairs, such that these conceptions satisfy CAS (and thereby satisfy CPC*). We have seen that this amounts to forming conceptions of things that satisfy SSC and ICC. I will, then, show that the following method can form such conceptions by defending the following argument:

> 4. There are properties of actual objects that are discovered by means of a Cartesian method M of forming conceptions of things.
>
> 5. All such properties are logically connected.
>
> 6. All such properties are also *essential* to such objects.
>
> 7. No other properties than the ones mentioned above are essential to such objects.
>
> 8. If (4)-(7), then M generates conceptions of actual objects that satisfy ICC and SSC (and so, satisfy CAS; thereby satisfying the antecedent of CPC*). (From the conjunction of 4-7, and the accounts of ICC, SSC and CAS)
>
> 9. Therefore, M generates conceptions of actual objects that satisfy ICC and SCC (and so, satisfy CAS; thereby satisfying the antecedent of CPC*). (From the conjunction of 4-7, and 8)

I promised that I would argue in this section that Descartes has a method for forming complete and accurate conceptions of the essences of objects. This was a bit misleading. For a *complete* defense of this claim will require some results from the next section as well. Therefore, in this section, I will argue for the truth of premises (4) and

(5). In the beginning of the next section I will finish the case for the Cartesian method by arguing for premises (6) and (7).

The Cartesian method of forming conceptions of things is nothing other than that of putting Descartes's theory of distinction into action. The method consists in determining the concepts that are transparently entailed by the concept of the principal attribute of a simple substance by applying TOD to the latter. For convenience, therefore, let's call the method, 'TODA' (for 'TOD applied'). With that said, let's illustrate TODA with some examples.

Consider my idea of a particular chunk of wax (call the chunk 'w'). In the first step of TODA, I reflect on w, and I clearly and distinctly perceive that its principal attribute (i.e., the only feature of w that is unchanging and constant under every kind of variation) is extension in length, breadth and depth. In the next step, I try to determine whether it is logically connected to any other concepts. It turns out that that concept is logically connected to several other concepts; for example, size, shape and position[123]; for an object that is extended (in length, breadth and depth) is transparently seen to have, necessarily, at least these other features. In the next step, I reflect upon these latter features, to see whether they, too, are logically connected to any other features. It turns out that each is transparently seen to logically connected to a disjunction of the concepts of all the coherently conceivable instances – or, as Descartes' would call them, 'modes' - of themselves. So, for example, to say that w has size entails that w has the size it

---

[123] This list of attributes is not meant to be exhaustive. For example, the capacities for motion and velocity logically supervene upon objects that are shaped and sized. Here and henceforth, let's just consider this sampling of attributes, for the sake of brevity.

189

actually has or some other determinate size or yet some other determinate size, etc.; similarly for the other features. Furthermore, it turns out that each such attribute is only rationally distinct from every other. Thus, [(s has size iff s has shape) iff s has position]. With the above conceptual work done, we are now in a position to make some deductions from it. This we do in the next step. Thus, we conclude that if w is extended, then w has shape, size and position. Further, we deduce that if w is extended, then w is spherical or w has a cube shape, or …and w is half a foot in diameter or w is 11 inches in length, or…and w is in my hand or w is on the mantle in a candlestick or… Finally, we may deduce that if w is extended, then w is spherical or w has a cube shape, or …iff w is half a foot in diameter or w is 11 inches in length, or…iff w is in my hand or w is on the mantle in a candlestick or… We may summarize these results concisely with the following bit of formalization. Let: Ex = w is extended; Sw = w has size; Shw = w has shape; Pw = w has position:

Step

Step	
1. Ew	Principal Attribute clearly and distinctly conceived
2. $Ew \leftrightarrow [(Pw \,\&\, Sw) \,\&\, Shw)]$	List of features transparently connected to it: non-principal Attributes
3. $[(Pw \leftrightarrow Sw) \leftrightarrow Shw]$	The relations between non-principal attributes: rationally distinct from one another
4. $Pw \leftrightarrow [(P1w \lor P2w) \lor P3w]$	List of features transparently connected to a non-principal attribute: modes
5. $Sw \leftrightarrow [(S1w \lor S2w) \lor S3w]$	List of features transparently connected to a non-principal attribute: modes
6. $Shw \leftrightarrow [(Sh1w \lor Sh2w) \lor Sh3w]$	List of features transparently connected to a non-principal attribute: modes
7. $Ew \leftrightarrow \{([[(P1w \lor P2w) \lor P3w] \leftrightarrow [(S1w \lor S2w) \lor S3w]) \leftrightarrow ([(Sh1w \lor Sh2w) \lor Sh3w]])\}$	Deduction from previous steps

TABLE A1: An Analysis of Finite Corporeal Substance Using TODA

190

With the completion of the previous step, we come to the end of an important stage of TODA. For we have now laid bare all of the concepts that are logically connected to the concept of a particular principal attribute of a simple corporeal substance. Furthermore, we have teased out an exhaustive account of how all of these concepts are logically interrelated.

Now in the next step, we need to set out all of the possible ways in which things may be distinct from one another. Now it turns out that the ways in which things may be distinct are materially equivalent to the ways in which the material implication relation can obtain between pairs of statements about a given entity. For statements about things being rationally distinct are materially equivalent to statements about distinct entities, A and B, predicated of a thing, x, such that Ax implies Bx and Bx implies Ax (call this '2-way implication'). Furthermore, statements about things being modally distinct are materially equivalent to statements about distinct entities, A and B, predicated of a thing, x, such that Ax implies Bx, but Bx doesn't imply Ax, or vice-versa (call this '1-way implication'). Finally, statements about things being really distinct are materially equivalent to statements about distinct entities, A and B, predicated of a thing, x, such that Ax doesn't imply Bx and Bx doesn't imply Ax (call this '0-way implication'). But this is just a mutually exclusive and jointly exhaustive list of the ways in which pairs of simple statements predicating something of an entity may stand in relations of material implication to one another: Necessarily, for any entity, x, features A and B, and pair of statements Ax and Bx, either:

(i) $\sim(Ax \to Bx) \mathrel{\&} \sim(Bx \to Ax)$ (which is equivalent to the *real* distinction)

191

(ii) $(Ax \rightarrow Bx)$ & $\sim(Bx \rightarrow Ax)$[124]   (which is equivalent to the *modal* distinction)

(iii) $(Ax \rightarrow Bx)$ & $(Bx \rightarrow Ax)$ (which is equivalent to the *rational* distinction)

It is important to note a crucial epistemological connection between the statements about entailment and the method of intelligible conceptual exclusion: the use of the latter rationally justifies the former. So, for example, the claim that $(Ax \rightarrow Bx)$ & $\sim(Bx \rightarrow Ax)$ is justified iff one can't intelligibly exclude x's A-ness from x's B-ness, but one can exclude x's B-ness from x's A-ness. I will argue later that this Cartesian insight is an important contribution to modal epistemology. With the distinction relations thus explicated, we have all we need to order the contents of Descartes's ontology into categories, based on the ways in which things may be distinct. This we do in the final step. Thus, from (2), we discover the set of non-principal attributes that follow necessarily and transparently from the principal attribute. From (3) and (iii)[125], we have the result that the non-principal attributes are only rationally distinct, since the possession of any one of them entails the possession of the others, and the lack of any one of them in w entails the lack of any of the others. From (4) and (ii), we have the result that position (or spatial location) is an attribute, and its various instances of being located in specific places are modes. For (4) entails that w can be spatially located, even if it isn't at location P1; but w wouldn't be at P1 if w weren't spatially located. Similar reasoning,

---

[124] Or vice-versa.

[125] Here and following, we will leave out all of the details of a full derivation, since it's clear that such derivations could be carried out.

192

when applied to premises (5)-(6), gets us similar conclusions with respect to shape, size, and the rest of the attributes and their instances.

Finally, (7) tells us that there is a deep unity between a principal attribute and all of its other attributes (and the varying modes though which these are expressed). This is because there is a mutual dependence between an instance of a principal attribute and its derivative attributes. Furthermore, the latter depend upon their being expressed as some mode or other (some mode or other from each non-principal attribute must obtain at any time in order for the principal attribute to obtain at that time). Anything lacking something displayed above can have no corporeal existence. However, as is displayed above, a thing need have no other feature in order to be a corporeal substance, since there are no entailment relations between a thing instantiating such features and any other feature of an any other entity (of the right sort[126]). Thus, such a structure is an absolutely *minimal* unit of corporeal reality. Therefore, we finally have an entity that's really distinct from all other things[127]: w's having extension does not entail that it has any other feature, and something's (including, what is impossible under the present assumption, w) having some other feature does not entail w's having extension. So by (7) and (i), we

---

[126] It would be nice if I could elaborate on the parenthetical qualification above. However, doing so adequately requires addressing apparent inadequacies in Descartes' criterion of substancehood, and I am not yet prepared to address them (beyond registering my conviction that they *can* be satisfactorily addressed in a way that is not incompatible with the claims of this paper). I hope to address these issues on another occasion.

[127] Here I mention only to set aside a complication raised by Descartes' theism, viz., that if no extended or finite thinking substance can exist apart from the originating or preserving power of God, then it looks as though such things are merely modally distinct from God.

193

have the result that the extended thing which is the wax is really distinct from all other things.[128]

Thus, with the completion of all the steps of TODA with respect to w, one has an exhaustive conception of a simple corporeal substance. For this conception includes (i) an exhaustive inventory of all of its constituent concepts, (ii) an exhaustive account of the how all of these constituents are logically interrelated, and consequently (iii) a complete ontological categorization of the thing conceived and its constituents. Furthermore, such a conception has neither one constituent more, nor one constituent less, than is required for the conception to be intelligible.[129]

---

[128] What about motion and rest? I haven't included such features in the discussion above because I haven't found an account of them that fully satisfies me. For on the one hand, the relationships between Motion and particular motions look to be the same as those between attribute and mode as we've been using these terms. On the other hand, if we construe Motion as an attribute of *extended substance* (i.e., as an invariant feature of an object), then we get the absurd result that extended substances are of necessity incapable of rest! Perhaps we can account for motion and rest in extended substances with a *disjunctive* attribute -- the attribute of Motion or Rest. I'm not entirely happy with this awkward solution, but I think that *something* in the vicinity of this solution must be right. That is, I think there must be *some* attribute that's related to the particular motions and restings of extended substance. For the notion of extended substance is rendered unintelligible when we exclude from it both motion and rest. Thus, each extended substance is capable of being either in motion or at rest for as long as it exists. Furthermore, the concepts of motion and rest are asymmetrically logically connected to concepts of particular motions and restings: excludability of the concept of motion renders conceptions of particular motions unintelligible, but not vice-versa. The same goes for rest. In short, motions and restings exhibit the same sort of phenomena that are true of those that hold between other attributes and modes connected with the principal attribute of extension, and this leads me to think that there is some attribute of extended substance that must account for this.

[129] I would add that it appears, at least initially, that a referent of such a conception has not one feature more, nor one feature less, than is absolutely required for it to qualify as a corporeal substance.

194

The previous discussion shows that (i) TODA is a method that generates conceptions of things, and (ii) that the constituent concepts of these conceptions are all logically connected in the requisite way. But this isn't yet enough. For recall that CAS also requires that (iii) we can see *that* the constituent concepts are logically connected. But it's not difficult to see that (iii) can be met. For all of the constituent concepts of a conception of Cartesian simple substance are connected by unavoidably obvious two-way conceptual entailment relations[130]. Thus, if one conceives of any of the attributes of such a substance, one cannot fail to notice each of its other attributes. So, for example, consider our earlier example of Cartesian simple *extended* substance. As we have seen, each constituent of its composite concept stands in unavoidably obvious two-way entailment relations to every other. For example, shape only applies to things whose particular shape is located somewhere or other; conversely, anything that occupies a spatial position must have shape. And no one who has run through an instance of TODA with respect to the concept of an extended substance, and who has conceived of a particular shape of a thing, could fail to conceive of that shape existing without occupying some spatial position or other, and conversely. Similar points can be made with respect to size and the rest of the attributes of extended substance. But if so, then since such conceptions are just those of simple substances generated by means of TODA,

---

[130] x stands in unavoidably obvious two-way entailment relations to y =def. (i) x stands in two-way entailment relations to y, and (ii) for any (properly functioning, adult) person S, if S conceives of x (at least, if S does so by means of TODA), then S cannot fail to (a) conceive of y and (b) conceive of x as standing in two-way entailment relations to y. The fact that two (or more) entities stand in unavoidably obvious two-way entailment relations seems to have an *epistemological* basis: each is rendered *unintelligible* if excluded from the others.

195

it follows that TODA is a method of generating conceptions that are not only logically connected, but such that we can't fail to see that they are. And if this is so, then we have completed our case for showing that TODA is a method that satisfies our three desiderata (i.e., (i)-(iii) above): TODA is a method that generates conceptions of things, the constituent concepts of these conceptions are all logically connected in the requisite way, and we are guaranteed to see *that* they are logically connected. This, therefore, completes our case for (4) and (5). I am also inclined to think that we already have at least *some* reason for accepting (6): that the properties thereby conceived are also *essential* to the things conceived. After all, how can we deny that (for example), macro-sized material objects are extended, shaped, sized, and located? I think it's clear, therefore, that we are already *prima facie* justified in accepting (6). However, I hope to say a little more on behalf of (6) (and (7)) (with reasons that are distinctively Cartesian in spirit) in the next section.

6. Defending (iii): Objects with Graspable Essences

It's time to complete the argument that Descartes has a method that generates conceptions of things that satisfy SSC and ICC (and so, satisfy CAS, thereby satisfying the antecedent of CPC*). The task remaining is, recall, arguing for two final premises of the argument: that (6) the properties discovered via TODA are *essential* to the objects conceived, and that (7) *only* such concepts are essential to such objects. I will give two Cartesian arguments for these premises: one *a priori* and one (at least partly) *a posteriori*. Each of the two premises receives support from both arguments.

196

The *a priori* argument is based on the goodness of God, whose existence is taken by Descartes to have been demonstrated. The argument can be stated as follows: If a *good* God is considering actualizing some possible world containing self-conscious creatures, then His goodness will prevent Him from actualizing a world in which such creatures are unable, even after their best efforts, to accurately conceptualize the natures of the objects in the world in which they exist – either by having no idea of what their natures are like, or by being deceived by their own innate phenomenology (e.g., after forming a faulty conception of a thing, the conception is attended with an irresistible conviction that the conception captures the essence of the thing.). So if He actualizes a world with self-conscious creatures, He will only actualize one in which they can form accurate conceptions of their natures. And His goodness will ensure that such creatures realize when their conceptions of things are accurate. He will attend such conceptions with an unmistakable indicator of accuracy, viz., indubitability. Therefore, since *we* are self-conscious creatures, then it follows that *our* world is one in which we can form accurate conceptions of things. Now it is only possible for us to discern, indubitably, exactly those properties of objects that we have mentioned above in relation to our discussion of TODA. Therefore, by the goodness of God, we can know that such properties are essential, and that objects have *only* such essential properties.

The *a posteriori* justification for (6) and (7) is basically an argument from theoretical parsimony: all other material phenomena are ultimately explainable via simple extended substances. So, for example, hardness, softness, rigidity, flexibility, being breakable, malleability, ductility, elasticity, sectility and all other physical properties

ultimately logically supervene on the properties of collections of suitably arranged, simple Cartesian substances. As Descartes states in *Principles of Philosophy*,

> "Consider how amazing are the properties of magnets and of fire, and how different they are from the properties we commonly observe in other bodies...*I have deduced the causes - which I believe to be quite evident - of these and many other phenomena from principles which are known to all and admitted by all, namely, shape, size, position and motion of particles of matter.* And anyone who considers all this will readily be convinced that there are no powers in stones and plants that are so mysterious, and no marvels ...that are so astonishing, that they cannot be explained in this way. In short, there is nothing in nature (nothing, that is, which should be referred to purely corporeal causes, i.e., those devoid of thought and mind) which is incapable of being deductively explained on the basis of these selfsame principles"[131].

The argument implied by the quote can be expressed as follows:

1. All observed physical phenomena can be explained via an account of matter, such that it has only the properties of size, shape, position, etc.

2. This account is the simplest that also explains all of the phenomena.

3. The account that is the simplest and also explains all the phenomena is the most rationally acceptable account.

4. Therefore, the account of matter that posits only the properties of size, shape, position, etc. is the most rationally acceptable account.

In short, the *a priori* argument tells us that a good God wouldn't let his self-conscious creatures go wrong about our best possible attempts at understanding the nature of reality. But since our best efforts at conceptualizing the nature of reality yield the conceptions got via TODA, then such conceptions accurately capture all and only the essential properties of such objects. And the *a posteriori* argument from parsimony tells us that since all

---

[131] CSM I, p. 279. Emphasis mine.

observed physical phenomena can ultimately be explained by simple substances with just the properties yielded via TODA, we have excellent reason to think that they have only such properties. But these results are nothing other than premises (6) and (7). Therefore, the case for the argument is complete: TODA is a method that generates concepts of objects that satisfy SSC and ICC – if there *are* any such objects. But if so, then they can satisfy CAS, and thereby satisfy the antecedent of CPC*.

It is only a short step from the immediately preceding discussion to showing that there are actual *objects,* such that we can form accurate conceptions of them, and such that such conceptions satisfy SSC and ICC. For in Section V, we formed conceptions of objects, (a) the former of which satisfy SSC and ICC (when the conceptions accurately represent their referents), and (b) the latter of which we *take* to be real. But if so, then the *a priori* argument from the goodness of God and *the a posteriori* argument from theoretical parsimony give us reason to think that such conceptions *are* accurate, and therefore that their referents are real. Therefore, we have reached the end of our long and arduous argument: Descartes has an ontology, such that the conceptions of some of its members satisfy SSC and ICC.[132] Therefore, since we have seen that some conceptions

---

[132] Despite what has been argued so far, one may still think that it is possible for there to be simple substances that have features that are logically independent of all the other features of such substances. That is, they neither entail nor are entailed by any feature (or sum of features) within the substance to which they belong. But if so, then I haven't yet shown that Cartesian simple substances satisfy SST and ICT.

In response to this, I would argue that there *couldn't* be essential features of a thing that don't stand in entailment relations to other features of that thing. For suppose for *reductio* that there *could* be a logically independent feature, F, of some arbitrary Cartesian simple substance, c. Then (by hypothesis) F would neither entail, nor be entailed by, c. But then (by the material equivalence that holds between implication relations and distinction relations, as I argued above) F would be really distinct from c.

satisfy SSC and ICC, it follows that such conceptions satisfy (CAS). And if so, then they can satisfy the antecedent of CPC*. And finally, since this is so, it appears that such conceptions can be used in connection with TOD to reach fully justified *de re* modal beliefs.

6. Evaluation of the Project

This is about as far as we can go in defending Descartes' CPC* as it currently stands. Unfortunately, it looks as though it doesn't go far enough. There are three major classes of criticisms that can be leveled against our Cartesian account of CPC, each class corresponding to a clause in the central premise of the Master Argument: That some of the essential properties of base level **objects** in the actual world are Lawless, as we've defined that term; that the **criterion** (CAS) is trivially true (in the sense that nothing satisfies it), thus rendering CPC* trivially true; and that the **method** of generating accurate conceptions of actual objects is problematic as it stands. As we will see, the last two classes of criticisms follow from the first.

6.1 Criticisms of Descartes' Ontology

First, it looks as though contemporary science has proved Descartes wrong about the nature of base-level entities. For whatever one wants to say about *other* possible worlds, his base-level extended substances don't correspond to fundamental physical entities in the *actual* world. For the latter have, in addition to those that Descartes lists, *other*

But then F would either be (a) a substance in its own right, or (b) a feature of some other substance, which contradicts our original assumption that F is a feature of c. Thus, there can be no such essential features of a thing. Therefore, I maintain that Cartesian simple substances satisfy SST and ICT.

200

essential base-level properties - e.g., *charge* - and these aren't clearly logically connected to the other essential properties. Indeed, charge seems to be a *Lawless* Feature of base-level entities in the actual world: one can conceptually exclude charge – apparently an *essential* property of particles -- from one's conception of Cartesian extended simple objects without thereby rendering the conception unintelligible. But if so, then Descartes' ontological picture is such that our accurate conceptions of them don't satisfy CAS, for they have Lawless Features.

6.2 Criticisms of the criterion (CAS):

The criticisms that apply to CAS follow from problems with Descartes' ontology. Of course, it would be an excellent criterion if the scientific picture Descartes had in mind were true: a world that is composed of entities whose essential properties are all capable of being incorrigibly, *a priori* teased out via reflection upon their most obvious (seemingly) essential property – extension – and the features to which it is transparently and necessarily related. Unfortunately, we now know that such a picture of reality is seriously inadequate. The essential properties of simple material objects aren't such that their adequate conceptions are all cognitively accessible and logically interconnected (at least not in the strong senses of these terms that we have been using). Rather, their essences have at least *some* Lawless Features. But if so, then Descartes' CPC* has once again fallen into the clutches of problem (2) mentioned early on in this essay: CPC* now looks to be trivially true, in that nothing satisfies its antecedent (since nothing satisfies CAS).

6.3 Criticisms of the method (TODA):

The criticisms applicable to TODA follow from problems with the objects as well. First,

the *a priori* argument from the goodness of God to the conclusion that we can grasp the

whole essence of an object is extremely problematic at best. For one thing, and as is well

known, Descartes' arguments for the existence of a good God are problematic. And if we

don't have other good reasons for this proposition, or it isn't otherwise *prima facie*

justified, then the *a priori* argument fails.

Second, as for the *a posteriori* argument: the *kind* of reasoning is plausible, but

the specific conclusion is false. As we have mentioned above, objects (whether

supervenient or base) have more essential properties than are dreamt of in Descartes'

philosophy. We now know that all the phenomena can't be explained by just those

essential properties (recall the "charge" example).

7. Summary and Conclusion

In summary, then, we've taken a careful look at Descartes' modal epistemology.

Spelling it out revealed that the most difficult issue facing his account concerns justified

modal inferences about actually existing objects -- in particular, the issue of ensuring that

the conception of the actual object is accurate. We then saw that his account of modal

epistemology can adequately handle this issue on certain assumptions – especially the

assumption that corporeal objects have a structure of Shallow and Lawful features.

However, while such an account of objects may have held promise in Descartes' day, it

202

doesn't in our own. Consequently, Descartes' modal epistemology is of extremely

limited use at best – at least in the form bequeathed to us.

Appendix II:
VAN INWAGEN'S CASE FOR MODAL SKEPTICISM

It's tempting to identify Van Inwagen's case for Remoteness Modal Skepticism with his 1998 paper, "Modal Epistemology" [133] (*ME*). However, to do so would be to miss a more comprehensive and compelling case, which extends into five other texts: "Ontological Arguments"[134] (*OA*) (1977), "Review of Richard Swinburne's *The Coherence of Theism*"[135] (*RS*) (1979), "The Problem of Evil, the Problem of Air, and the Problem of Silence"[136] (*EAS*) (1991), *Metaphysics*[137] (*M*) (1993), and "Materialism and the Psychological Continuity Account of Personal Identity"[138] (*MPI*) (1997). In what follows, I will use the relevant material that can be gleaned from these six texts to offer a systematic case for Remoteness Modal Skepticism. Toward that end, I will first lay out

[133] *Philosophical Studies* 92 (1998), pp. 67-84. Reprinted in *Ontology, Identity, and Modality: Essays in Metaphysics* (Cambridge UP, 2001), ch. 10.

[134] *Nous* 11 (1977), pp. 375-95. Reprinted in *God, Knowledge, and Mystery: Essays in Philosophical Theology* (Ithaca and London), pp. 22-41.

[135] *The Philosophical Review* 88 (1979), pp. 668-72.

[136] *Philosophical Perspectives, Vol. 5: Philosophy of Religion* (1991), pp. 135-165. Reprinted in *God, Knowledge, and Mystery: Essays in Philosophical Theology* (Cornell UP, 1995), pp. 66-95.

[137] There are two editions, both from Westview Press: 1993 and 2002. All citations refer to the 2002 edition.

[138] *Philosophical Perspectives 11: Mind, Causation, and World* (1997), pp. 309-17. Reprinted in *Ontology, Identity, and Modality: Essays in Metaphysics* (Cambridge UP, 2001), ch. 9. All citations are from the latter.

204

the various accounts of modal justification Van Inwagen discusses in the aforementioned

texts. I will then explain the several criticisms he offers within them, and how they apply

to each type of account of modal justification. Finally, I will bring all these points

together to show how they provide a powerful case for Remoteness Modal Skepticism.[139]

1. Varieties of Modal Justification

Van Inwagen considers and evaluates at least six different accounts of knowledge or

justified belief regarding modal claims. These can be divided into *Imaginability*

accounts and *Argument* accounts. Let's look at each of these in turn.

1.1 Imaginability Accounts

Imaginability accounts vary widely, but they all have in common the idea that mental

representations of various sorts provide a fundamental source of justification for modal

claims. These mental representations may be propositional, pictorial (in the sense of a

mental image), or a mix of both. But, whichever account of mental representation is

endorsed, the fundamental intuition driving such accounts is that mental representations

are capable of *accurately depicting* ways the world must be, can be, or could have been,

and this in turn legitimates their role in justifying modal claims. Imaginability accounts

can be divided up into three classes -- *Liberal*, *Moderate*, and *Conservative* – depending

---

[139] It should be noted that in addition to van Inwagen's arguments discussed below, he frequently appeals to George Seddon's modal skeptical arguments given in his "Logical Possibility", *Mind* 81 (1972), pp. 481-94. But since we've covered Seddon's main argument in chapter 1, we need not rehearse them here. I am indebted to van Inwagen for his reference to Seddon's article in his writings.

on the stringency of the constraints they put on justification-conferring imaginings. Let's look briefly at each class.

Liberal Imaginability accounts have at least the following three general characteristics. First, the relevant sort of justification-conferring representation is *propositional* representation: constructing a set of propositions that represent a possible world or state of affairs. Second, single, simple propositions are often sufficient, and indeed typical, for justifying modal claims; and third, the justification-conferring properties of such propositions are explained solely in terms of "analytic" considerations, such as the laws of logic and the meanings of terms. Van Inwagen discusses and critiques at least three versions of Liberal Imaginability: *Logical Possibility*, *Presumption of Possibility*, and *Empirical Propositions*.

The most unconstrained account of the Liberal sort is the *Logical Possibility* account. According to this account, if a possibility claim doesn't entail a contradiction, then its referent is metaphysically possible.[140] On this account, then, if 'I exist and nothing material exists' doesn't entail a contradiction, then there is a metaphysically possible world at which I exist and nothing material exists.

Second, there is the *Empirical Propositions* account.[141] An "empirical proposition" is one whose truth-value doesn't depend on the meanings of words, but can only be shown true or false via observation. According to this account, then, all empirical propositions are metaphysically possible. Thus, golden mountains are possible

---

[140] Van Inwagen discusses this account in "Modal Epistemology", pp. 247-248.

[141] "Ontological Arguments", pp. 33-34.

206

because the truth-value of 'There is a golden mountain' can't be determined by just understanding it or reflecting on the meanings of its words and their interrelationships, but only by empirical investigation. It's hard to see whether there's a material difference between the Empirical Propositions and Logical Possibility accounts, but since Van Inwagen treats them separately, I'll go along with his distinction here -- at least for now.

Lastly, there is the *Presumption of Possibility* account, which is slightly weaker than the previous two; it makes concessions to our lack of logical and conceptual omniscience. According to Presumption of Possibility, a possibility claim is to be presumed true unless proven false.[142] So on this account, 'I exist and nothing material exists' may or may not entail a contradiction. But if it does, it doesn't wear it on its sleeve. One is thus justified in taking its referent object or state of affairs to be metaphysically possible unless or until proven false.

Moderate Imaginability accounts have the constraint of Liberal accounts (actual or apparent lack of surface-level incoherence), but go beyond it by adding that such imaginings must include at least some fleshing out in terms of a *scenario* in which the modal proposition is true, where the imagined scenario consists of pictorial mental representations, propositional representations, or a combination of both. Presumably, the rationale for this constraint is that the more complete the representation is, the less likely it is that one has overlooked an inconsistency entailed by the target modal claim. Richard Swinburne's account is a paradigm case of a Moderate Imaginability account. Here is one an example of a justification-conferring imagining according to Swinburne:

---

[142] Van Inwagen discusses this account in *Metaphysics*, 2nd ed., pp. 107-108.

207

"Imagine ...your thinking being equally coherent however men mess about with your brain. Imagine too that you cease to feel any pains, aches, and thrills, although you remain aware of what's going on in...your body. You gradually find yourself aware of what is going on in bodies other than your own and other material objects at any place in space...an ability which proves unaffected by men interfering with lines of communication...You also come to see things from any point of view you choose...You...find yourself able to move directly anything which you choose, including the hands of other people...You also find yourself able to utter words which can be heard anywhere, without moving any material objects. However, although you find yourself gaining these strange powers, you remain otherwise the same ...Surely anyone can thus conceive of himself becoming an omnipresent spirit."[143]

Thus, according to Swinburne, since one can tell a coherent story like this about what it's like to become an omnipresent spirit, one is justified in taking 'Possibly, I become an omnipresent spirit' to be metaphysically possible, as well as, 'Possibly, there is an omnipresent spirit', which is entailed by it.

Conservative Imaginability accounts typically endorse the constraints of Moderate accounts (surface-level logical consistency, plus imagining a coherent scenario according to which the target possibility claim is true). But the individuating constraint of Conservative account is that it must *seem* or *appear* to one that the target possibility claim *P* is true in a possible world, where this appearance moves one to believe that *P*. Stephen Yablo calls this constraint, "the Modal Appearance Test".[144]

---

[143] *The Coherence of Theism* (Oxford: Clarendon, 1977, p. 105. Quoted from Van Inwagen, "Review of The Coherence of Theism", *The Philosophical Review* (1977), p. 670-671.

[144] "Is Conceivability a Guide to Possibility?", *Philosophy and Phenomenological Research* 53 (1993), pp. 1-42, esp. pp. 4-7. Two other philosophers who hold to the "appears possible" requirement – thus making them Conservatives -- are James Van Cleve and Paul Tidman. Van Cleve ("Conceivability and the Cartesian Argument for Dualism", *Pacific Philosophical Quarterly* 64 (1983), pp. 35-45) distinguishes between

The motivation for the constraint seems to be similar to the one for the Moderate constraint. For just as Moderates worry that the Liberal assumption that inability to find a contradiction in a single, simple *proposition* is sufficient reason to think it possibly true, so Conservatives worry that the Moderate assumption that inability to find an incoherence in a *scenario* is sufficient reason to think its target proposition is possibly true. In short, there is a worry for both Liberal and Moderate accounts about whether the *inability* to see that P is *impossible* is sufficient evidence to think that P is *possible*. Conservative accounts attempt to redress these worries by requiring that it must also *appear* to one that P is possible. As we saw in chapter 2, Yablo's account is a textbook Conservative account.

1.2 The Argument Account: Leibnizian Possibility Proofs (OA, M)

The final sort of modal epistemology Van Inwagen considers isn't an imagination account at all, but rather a means of demonstrating that something is metaphysically possible via deductive proof.[145] Van Inwagen has in mind here Leibniz' method of *arguing* for possibility claims.

Leibniz' method depends crucially on a metaphysical thesis about properties, viz., that any complex of simple positive properties is possible. According to Leibniz, a property is *simple* iff it has no other properties as constituents. It's hard to think of cases

what he calls "weak conceivability" and "strong conceivability", where the possibility that P is *weakly conceivable* iff one can intuit that P is impossible, and the possibility of P is *strongly conceivable* iff one can intuit that P is possible. According to Van Cleve, only strong conceivability is a reliable guide to metaphysical possibility. Tidman adopts Van Cleve's strong/weak conceivability distinction in "Conceivability as a Test for Possibility", *American Philosophical Quarterly* 31 (1994), pp. 297-309.

[145] The following is based on Van Inwagen's discussion in *Metaphysics*, pp. 108-109.

of simple properties, but let's suppose, for the sake of discussion, that *being a color* is a simple property. And a property is *positive* iff it's not a negation of another property. Thus, *being a white male* is a positive property; but *being a non-cat* is not, as it's the negation of the property *being a cat*.

Leibniz' argument for his thesis is that the only way for two or more properties to be incompatible with one another is if one is the negation of the other. But since no positive property is a negation of another property, it's impossible for a complex consisting solely of positive properties to be incompatible. Now one might grant that any positive property is compatible with any other positive property, but worry that such a property might have a negative property as a *constituent*, and *it* might be incompatible with one of the other constituent positive properties. This is where the qualification about simple properties comes in: So long as all the *simple* properties are positive, there can be no chance of incompatibility within a complex of properties. And since any complex of compatible properties is possible, his thesis follows: all complexes of simple positive properties are possible. Thus, suppose *being round*, *being red*, and *being a thing* are all simple positive properties. Then the properties are compatible, and thus being a round, red thing is possible; in which case, "Possibly, there is a round, red thing" is true.

Leibniz thought the scope of this method's justification-conferring power is virtually boundless, and thus capable of justifying High Modal claims. Most famously, he used this method to prove the key premise of the modal ontological argument -- that a perfect being is possible – on the grounds that a perfect being's essence consists solely of

210

simple positive properties, as all perfections are, or are composed of, simple, positive properties.

These, then, are the accounts of modal epistemology that Van Inwagen discusses. He concludes that they are all inadequate, and thus unable to move us beyond Remoteness Modal Skepticism. But what *reasons* does he offer for rejecting these accounts? This is the topic of the next section.

2. A Catalogue of Criticisms

2.1 Undecidable Armchair Claims: (OA)

It's commonly thought that conceptual issues or claims must be tractable or decidable, and decidable via armchair methods, such as logical and mathematical reasoning, thought experiments, etc. Unfortunately, upon closer inspection we see that lots of "conceptual" claims are probably not decidable in this way. So, for example, consider the claim that there is a run of four sevens in the decimal expansion of π. This is a conceptual claim, and yet armchair methods and logical tools fail us when called upon to help determine its truth-value.[146] Thus, some conceptual claims may well be undecidable via armchair methods.

Undecidable Armchair Claims would seem to have worrisome implications for various imaginability accounts of modal justification. For if there is no *a priori* guarantee that one will be able to determine the truth-value of a conceptual, armchair claim, then the failure to find a contradiction within a proposition (or in relation to other

---

[146] "Ontological Arguments", p. 32.

211

propositions) is a questionable basis for thinking it to be possible. This is especially so for

Liberal accounts, as they take such an inability to be good evidence that the thing

imagined is possible.

2.2 Concept/Nature Gap: (OA, M).

Kripke and Putnam have shown that there is an epistemological gap between *concepts*,

which are the constituents of propositions, and *natures*, which partially constitute things

in the world.[147] For a concept is a constituent of the meaning of a word only if anyone

who understands the word is thereby able to grasp that concept. So, for example, when I

grasp the concept of *being a bachelor*, I thereby grasp the concept of *being unmarried.*

But in very many cases – including very many philosophically interesting cases -- one

can understand the meaning of a word without thereby grasping its referent's nature, as

the latter is often unrepresented in the former. In order to gain knowledge of the nature

of a thing, therefore, one must do *empirical research*, as armchair-based investigation of

one's concept is *inherently* incapable of yielding such knowledge.[148]

Concept/Nature Gap makes trouble for all three sorts of imaginability accounts of

modal epistemology Van Inwagen considers. For the justification-conferring power of

those accounts depends upon the capacity of an imagining to accurately represent modal

and non-modal facts about the world. But if there is a radical disconnect between

---

[147] Kripke, Saul. *Naming and Necessity* (Harvard UP, 1970); Putnam, Hilary. "The Meaning of 'Meaning'." *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science*, vol. 7, ed. Keith Gunderson (Minneapolis: University of Minnesota Press, 1975), pp. 131-193. Reprinted in *Mind, Language and Reality: Philosophical Papers*, vol. 2. (Cambridge UP, 1975), pp. 215-271.

[148] "Ontological Arguments", pp. 33-34.

concepts and natures, then this power is called into serious doubt.[149] This doubt cashes

out a little differently for each kind of imaginability account.

As for Liberal accounts, recall that all such accounts involve only propositional

imagining, and that the only constraint on such accounts is logical or conceptual

consistency. But Concept/Nature Gap shows this constraint to be insufficient. For

example, 'Hesperus is not Phosphorus' and 'Water is not H20' are logically and

conceptually consistent, despite the fact that they're necessarily false, due to

metaphysically necessary facts about the nature of the referents of these statements – the

identity of Hesperus and Phosphorus in the former, and the identity of water and H20 in

the latter. Further, the fact that Concept/Nature Gap is a *widespread* phenomenon rules

out as a plausible response the retreat to fallibilist versions of Logical Possibility, such as

Presumption of Possibility.

For Moderate accounts: Recall that Moderate accounts add to the sole constraint

of Liberal accounts – consistency -- the constraint that a modal claim be backed by a

fleshed-out scenario. But Concept/Nature Gap shows that the added constraint isn't

sufficient to confer *prima facie* justification on an imagining. For the criticism points out

that conceptual information gotten from the armchair often does not contain information

about the natures of things in the world. But since many modal claims crucially depend

on facts about the natures of their intended referents, there will be many cases in which a

---

[149] In recent years, several philosophers – most notably David Chalmers and Frank
Jackson -- have developed accounts of modal epistemology based on two-dimensional
semantics that attempt to handle the problem of Concept/Nature Gap. I've discussed this
account in chapter 3. I'll discuss some of the details of two-dimensionalist modal
epistemology in Appendix III.

scenario, no matter how detailed, is *inherently* incapable of helping us determine the truth-value of the relevant modal claim.

For Conservative accounts: Recall that such accounts add to the two constraints of Moderate accounts – consistency and backing by a scenario – a third condition of the Modal Appearance Test, which requires that the modal claim *appear* or *seem* true to you. If so, then Concept/Nature Gap gives rise to a dilemma: if an imagining based on armchair information makes it seem to one that a modal claim is true, then since such information leaves out facts about natures, there is good reason to doubt (or perhaps to consider inscrutable) the reliability of such seemings. On the other hand, if such imaginings fail to pass the modal appearance test, then the relevant modal claim is unjustified on the account's own terms. Therefore, either way, Concept/Nature Gap renders dubious the epistemic status of Conservative imaginings for High and Middling modal claims. Of course, there will be many imaginings that pass the modal appearance test and are justified, viz., those that aren't remote from ordinary experience – i.e., Low and (lower-) Middling imaginings. But then this makes no real advance over Van Inwagen's Remoteness modal skepticism.

In short, concepts, and thus propositions containing them, often lack information about the natures of things. And if so, then analysis of the meanings of words and their interrelations is a dubious guide for determining the modal status of facts in the world. Indeed, if concepts and natures can and do come apart in this way, then even if one were logically omniscient, and thus knew all the conceptual and logical entailments of and between the concepts of a proposition, it may yet well be that one is *still* unable to

214

determine the proposition's true modal status. This criticism applies to all versions of imaginability mentioned above, as they all typically appeal to conceptual knowledge gotten from the armchair.

At this point, one might object that while Concept/Nature Gap certainly poses problems for *propositional* construals of Moderate and Conservative imaginability, it leaves *pictorial* accounts untouched. Furthermore, one might think that while fleshed-out scenarios may not be able to shed light on the nature of some actual-world referent, this isn't a problem in many interesting cases. For many modal claims only require that there is *some possible nature or other* that has the imagined features. But surely the fleshed-out scenarios involved in Moderate and Conservative accounts provide sufficient evidence for the possibility of such natures. For if one can imagine a scenario involving some of the features captured by my armchair concept, then there must be *some* possible entity or other that could instantiate those features. This reply is addressed by another of Van Inwagen's criticisms of imaginability accounts of modal epistemology, to be discussed next.

2.3 Weak Deep Structure Doubts (Deep Structure Doubts$_w$): (*OA*, RS, MPI, M, ME) When it comes to High and Middling possibility claims, there are often significant worries that what one is imagining may well be metaphysically impossible, due to some deep incompatibility among the properties required to make the imagined referent actualizable. Call worries of this sort, *Deep Structure Doubts*.[150] As will become clear in

---

[150] Van Inwagen makes frequent reference to George Seddon's paper, "Logical Possibility" (*Mind* 81 (1972), pp. 481-94) as the basis of the objection I am here calling 'Deep Structure Doubts'. In that paper, Seddon argues (among other things) that there are

215

a moment, Deep Structure Doubts pose a problem for all imaginability accounts on our list.

Arguably, Van Inwagen has two different versions of this criticism – a weak version (henceforth 'Deep Structure Doubts$_w$') and a strong version (henceforth 'Deep Structure Doubts$_s$'). According to *Deep Structure Doubts$_w$*, there is at least a *prima facie* basis for thinking that no metaphysically possible set of inter-compatible properties could undergird *a particular individual* or *kind of individual* referred to in the relevant thought experiment. Van Inwagen gives a large number of cases in support this claim: cases involving claims about the possibility of transparent iron[151], a human being drinking lots of alcohol (relative to one's height and weight) without it affecting their sobriety, a moon made out of cheese[152], science-fiction brain-state-transfer cases of various sorts discussed in the personal identity literature[153], and naturally purple cows[154]. The basic idea is that the more a thought experiment envisages relatively significant changes in the structure of

---

lots of modal claims confidently trotted out by philosophers about individuals and kinds that, although logically consistent, are nonetheless probably metaphysically impossible. The reason is that the envisaged scenarios require a change within the individuals or kinds so significant that it is no longer *the same* kind or individual. So, for example, the claim that there are carnivorous rabbits is not self-contradictory, but for all that, carnivorous rabbits are probably metaphysically impossible. For one would have to change the laws of nature or the structure of the rabbit so drastically that it would no longer be *that* rabbit – or even *a* rabbit.

[151] "Ontological Arguments", pp. 32-33.

[152] Both cases are in "Review of Richard Swinburne's *The Coherence of Theism*", pp. 671-672.

[153] "Materialism and Personal Identity", pp. 146-148.

[154] "Modal Epistemology", pp. 254-255.

an individual or kind of individual, the less defensible it is to say that you're still envisaging the *same* individual or the same *kind* of individual; the likelihood that you're envisioning a metaphysically possible state of affairs involving *that very individual*, or an individual of *that very kind*, is low, or at best, inscrutable. But if so, then what one is envisioning doesn't justify the corresponding modal claim. It's worth looking at a couple of Van Inwagen's cases in some detail to feel the force of the argument. The first one, taken from OA, involves a thought experiment intended to support a modal claim about the possibility of transparent sheets of iron. Here it is:

> "Consider the proposition that some three-inch-think sheets of iron are transparent to visible light...Is it possibly true? I don't know and neither does anyone else...Perhaps if the charge on the electron were not 1.6×10^{-19} coulombs but, say, two-thirds that, the crystalline structure of iron would be altered in such a way that a beam of photons could pass through a sheet of iron without melting or vaporizing it. But, for all I know...the electronic structure of the iron atom is such that, no matter what the charge on the electron was, iron atoms would have to fit together in such a way as to bar the passage of radiation in the visible frequencies...I think, therefore, that no one has any good reason to believe that the proposition

> Some three-inch-thick sheets of iron are transparent to visible light.

Is possibly true."[155]

The second, which occurs in MPI, is in response to a thought experiment given by Sydney Shoemaker in support of a version of psychological continuity accounts of personal identity. Shoemaker's thought experiment involves transferring a "person" from one brain to another via the transfer of their global brain-state, or all the information stored in their brain at that time, from one brain to another. In reply, Van Inwagen raises

---

[155] Pp. 32-33

217

the point that what's possible for one kind of thing may not be possible for another.[156]

Thus, if we don't know the nature of the human brain and consciousness, then we can't justifiably know that a "computer disk" model represents the way the human brain stores information, as opposed to an alternative model such as, e.g., a "city" model of information storage. But if it's like the latter, then the sort of "transfer" imagined may well be metaphysically impossible. Unfortunately, armchair investigation can't help us determine *which* model is correct, in which case the epistemic status of the corresponding modal claim is inscrutable at best from the armchair.

2.4 Strong Deep Structure Doubts (Deep Structure Doubts$_S$) (MPI, ME):
The criticism of Deep Structure Doubts$_w$ was that it's implausible that or inscrutable whether there's a possible world at which *a particular individual* or particular *kind* of individual could possibly exist given the envisaged changes represented in the thought experiment; that is, holding a particular individual or kind fixed, it's doubtful or inscrutable that *that very individual or kind* could exist given the changes asserted in the modal claim. By contrast, the Deep Structure Doubts$_S$ criticism is that it's implausible or inscrutable whether there's a possible world at which *any* relevant individual or kind of individual exists and has the properties represented in the thought experiment. We can come at the idea from another direction by exploiting some ideas from Kripke and Putnam. Kripke pointed out that modal illusions can occur if one isn't careful to hold fixed the referent of a thought experiment, which can lead to a conflation of the referent with some entity qualitatively identical to, yet distinct from, the actual referent (whether

---

[156] *Ibid.*, pp. 146-147

218

the referent is an individual or natural kind).[157] Call any such entity an *epistemic counterpart*. What makes epistemic counterparts possible? This is where Putnam comes in. According to Putnam, we associate things with what he calls *stereotypes* – perceptual and behavioral features of a thing or kind learned from ordinary experience.[158] Unfortunately, the nature of a thing or kind and its stereotype can come apart – a nature could have a different stereotype, and a stereotype could be had by a different nature. In light of our discussion of epistemic counterparts and stereotypes, we can characterize Deep Structure Doubts$_s$ as the criticism that we're not justified in thinking that there is a possible world in which even an *epistemic counterpart* of the relevant individual or kind represented in the thought experiment exists. You may be able to envisage or represent a coherent set of inter-compatible "surface" or stereotypical properties that you take to be of a thing, all right, but it's doubtful whether any set of inter-compatible deep structure properties could undergird them. Thus, an undercutting or rebutting defeater for the possibility of carnivorous *rabbits* is a Deep Structure Doubt$_w$; an undercutting or rebutting defeater for the possibility of carnivorous *rabbit-like creatures* is a Deep Structure Doubt$_s$. Van Inwagen seems to be advancing a Deep Structure Doubt$_s$ criticism against high-flying, science-fiction thought experiments in MPI:

> "...I am similarly suspicious of all of the other imaginary technology that turns up in discussions of personal identity. The Star Trek transporter beam, for example, is suspect indeed. A one-second pulse of electromagnetic radiation that carried enough information to restructure a particular human organism...that arrived at a receiver a few meters across would simply vaporize the receiver and everything else in the vicinity and

---

[157] *Naming and Necessity*, Lecture III.

[158] "The Meaning of 'Meaning'", pp. 227-252. See esp. pp. 249-252.

219

everything it had happened to encounter on the way there. Similar considerations from the theory of signals show that it would take months or years to transfer the amount of information from one brain to another that is required by the brain-state-transfer machine; try to speed up the process significantly and you will just melt all those little wires attached to the shiny cap on the recipient's head."[159]

Thus, according to this case, serious doubts are raised as to whether *any* kind of brain – not just human brains – could undergo such a "transfer". The same thing holds for the transfer machine. It's not just that the Star Trek sort of machine couldn't work – couldn't play the "transfer" role -- but that no *possible* machine could do it. For such a transfer may violate not just *our* natural laws, but all *possible* natural laws regarding the energy required to transfer information – maybe, on *any* set of natural laws, the rapid transfer of information would require so much energy that the equipment would melt.[160] To the reply that one could grant the point about the metaphysical necessity of the laws about energy and information transfer, but alter the scenario so that the process goes slow enough to prevent "melting", Van Inwagen worries that the person/brain is changing during the process – it may be metaphysically necessary for any sort of brain – or perhaps even *brain-like-thing* – that it all needs to be "transferred" in a relatively short time – just as one can't paint a single painting on a series of canvasses, each one rapidly replaced by another. Another way to put the point about Deep Structure Doubt$_s$: think of roles and

---

[159] pp. 147-48

[160] Ibid: "...I am willing to bet that the statement, "For any given level of efficiency of information transfer, the greater the amount of information passing through that channel per unit of time, the greater the amount of energy passing through that channel per unit of time" is *a necessary truth*, a principle that is a consequence of *any* coherent set of laws of nature."

220

occupants. It's not just that the envisaged occupant couldn't play the role; it's that there is *no* possible occupant that could play the role.[161]

We can push Van Inwagen's Deep Structure Doubts$_s$ criticism further. For as it's currently stated, its force only applies on a case-by-case basis, since it provides no principled way to determine beforehand its force with respect a given imagining. However, it seems reasonable to think that our doubts about a viable undergirding structure for a given imagined stereotype is directly proportional to the degree of *dissimilarity* to the nature of the thing characterized by the stereotype in the *actual* world.[162] We may well be able to imagine a scenario according to which some alien substance, XYZ, has H20's stereotype. However, if XYZ is significantly dissimilar in nature to H20, then to that extent one's confidence that such a thing could undergird the water stereotype should be diminished. If this rationale is on the right track, then we can say that Deep Structure Doubts$_s$ undercut the force of High and (many) Middling Liberal and Moderate imaginings, respectively, on the grounds that their natures are significantly dissimilar to those bearing the relevant stereotypes in the actual world.

Many philosophers seem to assume that if they can imagine an array of inter-compatible stereotypical properties of a thing or state of affairs, then there must be some corresponding metaphysically viable deep structure – some set of inter-compatible

---

[161] If we are to speak more carefully, we should say that it depends on the modal claim in question. There are Strong Deep Structure Doubts about the possibility of XYZ playing the role of water, for instance. In this sort of case, the Deep Structure Doubt is that there may well be no *other* possible occupant for the role than the *actual* occupant.

[162] I develop the argument for this claim in more detail in the next chapter.

221

"micro-structural" properties -- capable of undergirding them. That is, many philosophers seem to accept the following principle (henceforth 'Roles Entail Occupants', or 'REO' for short):

> (REO): If one can imagine a stereotype, then there is at least one possible nature that instantiates that stereotype.

With Deep Structure Doubts, Van Inwagen undercuts REO. According to this criticism, then, a thought experiment may or may not involve a lot of detail. But even if it does, there are significant doubts about whether there is any inter-compatible or metaphysically possible "deep structure" capable of undergirding the "surfacy" details involved in the thought experiment.[163]

Deep Structure Doubts of both sorts make trouble for all versions of imaginability on our list. For Liberal accounts, the problem is that the target proposition may be a logically consistent proposition describing the stereotype of a candidate thing or kind, but there are significant doubts about a metaphysically possible underlying nature -- or at least a *relevant* nature -- capable of undergirding that stereotype.

---

[163] But mustn't there be *something* undergirding the stereotype in my imagining? After all, didn't Descartes teach us that the notion of an attribute without an underlying substance is unintelligible? I'm not so sure. For if Descartes is right, then it would be a *tour de force* against, say Descartes' evil demon hypothesis; *a fortiori* would it be so against Berkeleyan idealism. At any rate, if one has remaining worries about the ability to imagine a stereotype without an underlying nature, consider Putnam's "XYZ water" case again. We may think he's imagining a different and genuinely possible natural kind exhibiting the water stereotype. But for all we know, the seeming possibility may be due to our imagining H20, and merely *labeling* it as XYZ. See Peter Kung's "Imaginability as a Guide to Possibility" (ms.) for the epistemic troubles brought on by the use of "labels" in thought experiments.

222

Deep Structure Doubts pose similar problems for Moderate accounts. For again, such imagined scenarios are based on armchair knowledge, which often leaves out knowledge of the natures of things. Rather, they typically involve stereotypical features of things. But if so, then doubts arise about viable undergirding natures.

Finally, Deep Structure Doubts pose a similar sort of dilemma for Conservative accounts that we saw arise for them from Concept/Nature Gap: either imaginings subject to Deep Structure Doubts pass the modal appearance test or they don't. If they do, then we have reason to doubt the reliability of the modal appearance test; if they don't, then Conservative accounts rightly judge them unjustified. Either way, then, Deep Structure Doubts raise serious worries about justifying modal claims. As we saw earlier in our discussion of Concept/Nature Gap, Conservative imaginings get the right results regarding the epistemic status of Low and Middling claims. But again, this is no advance over Van Inwagen's Remoteness Modal Skepticism.

Notice that unlike Concept/Nature Gap, Deep Structure Doubts are just as much of a problem for pictorial imaginings as they are for propositional imaginings. For both sorts of imaginings, if based solely on armchair information, go no deeper than the ordinary observational and behavioral properties of things. As such, they don't touch the underlying natures of their referents. But if not, then, again, doubts about viable underlying natures arise.

2.5 Parity (M, ME)

At least when it comes to a given possibility claim remote from the practical concerns of daily life, the denial of such a claim typically has just as much going for it, epistemically,

as the original claim. If so, then each cancels out the epistemic force of the other. Call any criticism with this general form, 'Parity'. Van Inwagen discusses and employs Parity in two places: M and ME. In M, he raises the point with respect to the modal ontological argument, critiquing the Presumption of Possibility account of imaginability in the process.[164] Thus, consider the key premise of the modal ontological argument:

1. Possibly, a perfect being exists.

By 'perfect being', Van Inwagen means an Anselmian being of the sort discussed in relation to the ontological argument – a being that has all perfections -- including necessary existence – and has them essentially. So if one accepts (1) (so understood), then if one also accepts Axiom S5, then one should accept the conclusion that, necessarily, a perfect being exists. Now although some resist the argument by denying or suspending judgment about S5, most see (1) as the most controversial premise. Why should anyone accept it? Well, an adherent of Presumption of Possibility would say that the claim contains no explicit contradiction, and since this is so, then one should assume the possibility of a perfect being unless and until one can prove (1) to be false. This is where an application of Parity begins. For unfortunately for the proponent of the modal ontological argument, the following statement contains no explicit contradiction, either:

2. Possibly, a knowno exists.

Van Inwagen defines a "knowno" as a being who knows that there are no perfect beings. Now both (1) and (2) are High modal claims, and both seem to be on an epistemic par: neither is *prima facie* intrinsically impossible. But they can't *both* be simultaneously

---

[164] *Metaphysics*, pp. 107-108.

presumed possible, for each entails the impossibility of the other: if there is a perfect

being, then it exists in all possible worlds; if it exists in all possible worlds, then it's *true*

that it exists in all possible worlds; and if it's true that it exists in all possible worlds, then

no being can *know* that no being exists in all possible worlds, since 'know' is a success-

term. Conversely, if a knowno is possible, then there is a possible world at which he both

exists and knows there are no perfect beings. And if he *knows* it, then it's *true* in that

world that there are no perfect, necessary beings. And since (assuming Axiom S5)

what's necessary doesn't vary from world to world, it follows that there is no perfect

necessary being. Thus, (1) is true iff (2) is false. The moral is thus that Presumption of

Possibility is false, since it leads to absurdities of the sort just discussed.

In *M*, Van Inwagen's employs Parity to construct a *reductio* against the Liberal,

Presumption of Possibility account of imaginability. But in ME, his use of Parity is

different in at least two ways. First, he uses Parity to make a criticism that has much

wider scope, as it applies to all Liberal and Moderate accounts. Second, his use of Parity

doesn't involve a *reductio*. Rather, his point in ME on this score is that high-flying

modal claims are subject to Parity, in which case each cancels out any positive epistemic

status the other may have had.[165] Thus, in ME, Van Inwagen considers two other

arguments involving controversial modal claims, one of which is the conceivability

argument for substance dualism. As with the modal ontological argument, the

conceivability argument for dualism has a controversial modal premise:

(3) It's possible that I exist and nothing material exists.

---

[165] "Modal Epistemology", pp. 243-245.

225

Here, his point isn't just that each high-flying possibility claim is on the same epistemic footing as its negation, but the slightly more developed one that high-flying modal claims, when used in an *argument* to support a contentious conclusion, are such that the argument is subject to parity of reasoning by either a G.E. Moore Shift argument or an "inverted" version of the original argument (i.e., the possibility premise is replaced with the negation of the conclusion, and the conclusion is replaced with the negation of the possibility premise), with the result that each argument cancels out any epistemic force that the other may have had.[166] Thus, consider the following generic version of the conceivability argument for substance dualism:

A
1. It's possible that I exist and nothing material exists.
2. Whatever is material is essentially material.
3. Therefore, I am not a material thing.

This argument has at least one modal premise – a premise with a modal operator – viz., premise (1)[167]. Call any such argument a *modal argument*. Now while Van Inwagen thinks we're justified in believing lots of modal claims, he thinks that when it comes to modal claims significantly remote from ordinary experience, the philosopher who has antecedent doubts about the conclusion shouldn't (or at least needn't) accept the crucial

---

[166] *Ibid.*

[167] I say 'at least' because (2) is also a modal premise if essential properties are to be analysed in terms of possible worlds. However, some (e.g., Kit Fine) argue that modal analyses are inadequate. I discuss modal and non-modal analyses of essential properties in chapters 3 and 4.

modal premise.[168] In order to get clear on his reasons here, it will help if we first lay out

two other arguments whose conclusions conflict with A's:

B
1'. It is possible that I exist and nothing immaterial exists.
2'. Whatever is material is essentially material.
3'. Therefore, I am not an immaterial thing.

C
1''. I am a material being.
2. Whatever is material is essentially material.
3''. Therefore, it is not possible that I exist and nothing material exists.

B is what Van Inwagen calls an *inversion* of A, and C is of course a G.E. Moore Shift

against A.[169] Now Van Inwagen's claim is that for *any* modal argument with a modal

premise as remote from ordinary modal claims as the one in A, one can take at least one

of the B or C strategies against it, and whichever is taken, the resultant argument will be

at least as plausible as the original, A-form modal argument, in which case the force of

each cancels out that of the other. And if that's right, then no such modal argument

provides a good reason for its conclusion.

The Parity criticism applies fairly clearly to the various Liberal and Moderate

accounts of imaginability. For, first, consider the Liberal accounts. Now we've already

seen how Parity raises problems for Presumption of Possibility, but the problems it raises

for Liberal accounts don't end there. Thus, recall the constraint on Liberal accounts in

general. Their only constraint is lack of known contradiction. But then there are bound

---

[168] *Ibid.*

[169] *Ibid.* The sample arguments are Van Inwagen's.

227

to be many claims such that both they and their denials satisfy that condition, in which case they're vulnerable to Parity.

Second, Parity raises problems for Moderate accounts of imaginability. For while Moderate accounts add a further constraint on justification-conferring imaginings – the fleshing out of scenarios – it's clear that many possibility claims and their denials will satisfy Moderate constraints (e.g., the God vs. knowno case mentioned above), and so they, too, will fall prey to Parity.

2.6 Unexcluded Alternatives (OA, RS, ME)

Thought experiments for High and Middling modal claims equally support other modal claims, in which case they can't *prima facie* justify the ones intended. Call this criticism, *Unexcluded Alternatives*. It's helpful to compare this criticism to the traditional epistemological problem of skepticism regarding our beliefs about the external world. One can put the problem in terms of the inability to rule out alternatives: I don't know if my experiences are of my hands, since that would require that my evidence is sufficient to rule out alternatives to my belief that I have hands. But my evidence is qualitatively identical to the evidence I would have if, say, I were a brain in a vat hooked up to a computer that caused hand-experiences within me. Similarly, I don't know that my cat could survive the death of her body, since that would require that my evidence is sufficient to rule out alternatives to my belief that she could. But my evidence for that belief amounts to my imagining my cat dying, followed by a cat-shaped, Casper-the-ghost-like ectoplasm floating out of her body. But this evidence can't rule out, for example, the alternative that while my cat didn't survive the death of her body, God

228

created a distinct cat-shaped ectoplasm immediately following her death, and had it "float" out of her body.

There is a difference between the traditional skeptical argument and the modal skeptical argument. For in the former case, one might reply that one need not rule out *all* alternatives in order to have perceptual knowledge. Rather, one need only rule out all *relevant* alternatives, that is, all alternatives that are live possibilities for the context in question, based on one's background evidence. So, for example, while at the zoo, I can know that I see a zebra, and not a mule painted to look like a zebra, since my background evidence doesn't make it likely that there are painted mules in the zoo. This response isn't available to the modal dogmatist, however. For in this case we're talking about knowledge of what is *possible*, in which case the scope of relevant alternatives encompasses everything that we can't rule out as absolutely impossible. It would seem, then, that the problem of Unexcluded Alternatives is a much nastier problem for modal knowledge than it is for perceptual knowledge.

Van Inwagen gives several examples of the problem of Unexcluded Alternatives. According to the first: Suppose Hume's Indian potentate had a philosopher on staff, and suppose he tried to determine whether solid water is possible. Trying to imagine water turn from liquid to solid wouldn't show anything, since the imagining can't rule out relevant alternatives, such as that the liquid water was *replaced* by some non-water solid.[170] According to the second, suppose you imagined a group of scientists standing around a computer and shouting, "It's a counterexample to Goldbach's Conjecture!"

---

[170] "Ontological Arguments", pp. 32-33.

229

That wouldn't *prima facie* justify that claim that there is a possible world at which Goldbach's Conjecture is false, for the thought experiment is equally supportive of, for example, the alternative that there's a glitch in the computer program that caused it to derive the *mistaken* conclusion that Goldbach's Conjecture is false.[171] According to the third case (*ME*), suppose you imagined a scientist accepting the Nobel Prize to a cheering audience, who then thanks all those who helped him in his long and difficult journey to create transparent iron, and then holds up something that looks like a sheet of glass. That wouldn't *prima facie* justify the claim that there is a possible world in which there is transparent iron. For the thought experiment is equally supportive of other claims, such as that a group of jokester scientists have gotten together to enact a burlesque of a Nobel Prize ceremony, or that a group of scientists have conspired to fool the public that they've created transparent iron, etc. Therefore, no thought experiment *prima facie* justifies a modal claim if it can't rule out alternatives.[172]

Unexcluded Alternatives poses problems for both Moderate and Conservative accounts of imaginability. For it shows that fleshing out a coherent scenario according to which a claim is true isn't sufficient to justify a modal claim, at least in practice. For scenarios based solely on information gotten from the armchair are typically compatible with claims that don't entail the target modal claim. And if so, then (a) it is shown that the mere fleshing out of a coherent scenario isn't a sufficient condition for modal justification, in which case Moderate accounts are defeated, and (b) such scenarios lack

---

[171] "Review of Richard Swinburne's *The Coherence of Theism*", p. 671.

[172] "Modal Epistemology", pp. 255-257.

230

sufficient detail to convince one that the relevant claim is true in it, in which case they fail the modal appearance test, thus raising worries about the usefulness of Conservative accounts.

2.7 Criticism of Leibnizian Proofs of Possibility (OA, M)

Recall that Leibnizian proofs of possibility turned on the notion of simple positive properties. For all simple positive properties are compatible with one another, and so if one can show that some structure is composed of all and only simple positive properties, then one will thereby have shown that structure to be metaphysically possible. But Van Inwagen argues that the notion of a simple positive property is implausible. For consider the property *not having parts*. This property seems to be a *negative* property on Leibniz' account, being the negation of the property of having parts. But suppose we call the property of not having parts 'simplicity'. Then *non-simplicity* would be a negative property – the negation of the property of being simple. But wait! The property of being simple and the property of not having parts are the same property! But we've argued that it is both a positive and a negative property, depending on what we *call* the property. This suggests that the notion of a positive property isn't a property of things as they are in themselves, but merely a conventional property. But if so, then the notion of a simple positive property can't be used to prove possibility.[173]

---

[173] "Ontological Arguments", pp. 30-31. A slightly expanded version of this criticism can be found in *Metaphysics*, pp. 108-109.

3. Bigger Points

3.1 Superficiality, Explanatory Impotence, and Vacuity

We have seen that Van Inwagen raises a large number of problems for Liberal and

Moderate Imaginability accounts, and it seems that they have a common root, viz., *the*

candidate imaginings are superficial. Say that an imagining suffers from *superficiality* if

it either (i) lacks significant detail or fleshing out, or (ii) it has such detail, but it is gotten

from the armchair, and not via empirical investigation. Then at least the Liberal and

Moderate accounts of imaginability are superficial, and this superficiality, in turn, gives

rise to Undecidable Armchair Claims, Concept/Nature Gap, Deep Structure Doubts,

Parity, and Unexcluded Alternatives. Thus, there is a plausible *prima facie* case for

thinking that no imaginability account of modal epistemology will be adequate if it's also

a superficial account. And this point, in turn, supports a necessary condition for any

adequate non-skeptical modal epistemology: in slogan form, "No Justification Without

Explanation" (NJWE):

> (NJWE) A non-basic impossibility, possibility or necessity claim is
> justified only if it's accompanied by an explanation for *why* it's
> impossible, possible or necessary.[174]

---

[174] Van Inwagen seems to endorse something like NJWE in a number of places, but it's
most strongly implied in his discussions of what I have called 'Deep Structure Doubts'
and 'Unexcluded Alternatives' above. In each case, he indicates that such worries
undercut the relevant modal claims unless one has an explanatory account of how the
stereotypical features flow from their underlying natures. So, for example, in his
discussion of imagination-based accounts in relation to the above-mentioned "transparent
iron" case, he says: "Can we imagine a world in which there is transparent iron? Not
unless our imaginings take place at a level of structural detail comparable to that of the
imaginings of condensed-matter physicists who are trying to explain, say,
superconductivity." ("Modal Epistemology", p. 255). Something like NJWE also seems
to be behind his discussion of the "Indian philosopher" case regarding imagination-based
justifications for the claim about the possibility of solid water discussed earlier. As he

232

But if so, then there is pressure to look for an account of modal justification that won't allow for superficial imaginings to play the role of justifiers. And since we've seen that Liberal and Moderate imaginability accounts inherently involve superficial imaginings, we must look elsewhere.

The natural place to look next is among the Conservative Imaginability accounts, and Van Inwagen does just that in ME.[175] There, he considers Yablo's Conservative Imaginability account. Now it's not clear that Yablo's account stipulates that superficial imaginings are illegitimate; nevertheless, it doesn't entail that modal evidence *must* be superficial, either; so the account isn't ruled out as implausible on that basis. Unfortunately, though, it quickly becomes clear that when it comes to High and Middling modal claims, the requisite sorts of explanatory detail are beyond our reach from the armchair[176]. But if so, then Conservative Imaginability accounts like Yablo's fail for a different reason: they're *vacuous*. The account of modal justification may well be reliable, but, unfortunately, none of our imaginings meet its evidential standards, in which case the relevant class of modal claims is beyond the reach of justification.[177]

---

puts it in "Ontological Arguments": "If our philosopher knew what we, or the scientists among us, know about water and heat (if he knew about the atomic constitution of matter, the electronic properties of the water molecule, hydrogen bonding, van der Waals' forces, the kinetic nature of heat, and a host of other things, and knew the correct mathematical descriptions of the relations that hold among them), then he might be able to determine by calculation that water has various solid phases." (p. 33).

[175] Pp. 253-258.

[176] See fn. 35.

[177] "Modal Epistemology", pp. 253-257.

Thus, we have a new criticism of Imaginability accounts, viz., *Vacuity*. Unlike the other criticisms, though, this criticism builds off the others. For the reasoning behind Vacuity is that the previous criticisms are rooted in the superficiality of the relevant imaginings, which thereby justified the requirement of explanatory detail for modal claims embodied in NJWE. And this, in turn, when combined with our inadequate knowledge of such detail in the vast majority of interesting cases, justifies Vacuity. And finally, this shows that Conservative Imaginability accounts – our last bastion of hope for a plausible account of imaginability – are doomed.

3.2 Dubious Basis (ME):

As we've seen, Van Inwagen grants that we have lots of modal knowledge, and much of it is not mysterious. Thus, we have "analytic" modal knowledge, knowledge of possibilities inferred from actualities, knowledge of impossibilities inferred via *reductios*, etc. We may even have some basic "non-analytic" modal knowledge about humdrum modal claims; and we can sometimes extend such knowledge when we combine it with facts about how the world is put together. But when it comes to non-analytic possibility and necessity claims remote from the practical concerns of daily life, we're just fooling ourselves if we think we have knowledge or *prima facie* justified belief about them. Such false confidence is due to at least two factors: (i) believing Logical Possibility to be a guide to metaphysical possibility[178], and (ii) institutionalized pretense to having acquired High and Middling modal knowledge[179].

---

[178] *Ibid.*, pp. 247-248.

[179] *Ibid.*, p. 246.

We've seen a host of problems with the Logical Possibility account of modal epistemology. However, with Dubious Basis, Van Inwagen has a different criticism in mind. It's the simple yet common fallacious slide that's often made in Philosophy: We see that logical *inconsistency* is a reliable guide to metaphysical *impossibility,* and then fallaciously infer that therefore logical *consistency* must likewise be a reliable guide to metaphysical *possibility.* But this slide is due to a failure to see that logical possibility isn't a real kind of possibility that maps onto the mind-independent world. Van Inwagen illustrates this as follows. Suppose you had a map that listed a number of landmasses as inhabitable, and some as uninhabitable, *but which made no claim to completeness regarding the labeling of uninhabitable landmasses.* Suppose further that someone were to say that an island is *cartographically habitable* just in case it isn't marked 'uninhabitable' on the map. Clearly, in such a case, although cartographical uninhabitability is a reliable guide to "metaphysical uninhabitability", one wouldn't be justified in taking cartographical habitability as a guide to "metaphysical habitability"! Similarly, from the fact that logical impossibility is a reliable guide to metaphysical impossibility, one shouldn't infer that logical possibility is a reliable guide to logical possibility. For there is good reason to think that the set of logical impossibilities -- i.e., the set of statements shown to be logically contradictory -- isn't coextensive with the set of metaphysical impossibilities.[180]

The other point – the one about institutionalized pretense – is justified in virtue of two broad considerations. First, we've seen that accounts of modal epistemology are

---

[180] *Ibid.,* pp. 247-248.

subject to a host of serious criticisms, such as Concept/Nature Gap, Deep Structure Doubts, etc.

Second, "non-analytic" modal knowledge is mysterious. For unlike other categories of knowledge – e.g. perceptual, memorial, and testimonial knowledge -- we have *no idea* as to what are the sources of non-analytic modal knowledge that would make it possible.[181] We may *think* we know – viz., via our knowledge of logical possibility – but we've just seen why such thinking is misguided. Granted, our "analytic" modal knowledge isn't completely free of mystery, but at least we have a start on the answer here, viz., knowledge of concepts and logical, semantical, and mathematical rules. This isn't so when it comes to "non-analytic" modal knowledge. Thus, "Possibly, there are purple cows" is mysterious in a way that "Necessarily, all bachelors are unmarried" is not.

At this point, the force of Dubious Basis makes itself felt. For if philosophers routinely and confidently make High and Middling modal claims, and yet (i) all accounts of modal epistemology fall prey to the worries listed above, and (ii) "non-analytic" modal knowledge is mysterious in a way that other categories of knowledge are not, then it's reasonable to start looking for a *non-rational* basis for such confidence. And a plausible explanation of this sort is the Rortian explanation of a community of peers who let you get away with such claims. Van Inwagen has an illustration for this mechanism as well. Thus, consider again our ability to make distance judgments by sight. Within a certain range, our judgments on these matters are fallible, yet fairly accurate. However, when it

---

[181] *Ibid.*

236

comes to judgments about objects very far away – such as judgments about the distance of the moon from the Earth – our abilities in these matters fail us miserably. Well, we know this now. But what about people who lived in pre-scientific times? Humans commonly judged that they were able to "just see" that the moon was about 30 miles away. And clearly, the confidence about such false claims was due to their proven success with distance judgments involving matters relevant to daily life. And this, in turn, led to an environment in which their peers let them get away with confident assertions about judgments about the distance of the Moon based on sight. Similarly, our confident assertions of High and Middling modal claims is due to our proficiency in making accurate Low modal claims, and this, in turn, grounds our community's accepting atmosphere regarding grandiose modal claims.

These mechanisms, then – superficially plausible yet fallacious assumptions about logical possibility, and a widespread aptitude for making accurate Low modal judgments creating an atmosphere of confidence about judging High and Middling modal claims -- undercut our justification regarding Liberal and Moderate imaginings. They also undercut our confidence in High and Middling Conservative imaginings. For one might well confuse an appearance of possibility with a false sense of confidence gotten from the two factors just discussed.

4. Taking Stock: Van Inwagen's Overall Case

It is now time to take stock, step back, and combine Van Inwagen's scattered criticisms to construct a fairly comprehensive case for Remoteness Modal Skepticism. The

237

argument can be spelled out as follows: There are two broad sorts of accounts of justification for High and Middling modal claims: Argument accounts and Imaginability accounts. The only Argument account that's been developed is implausible, as it relies on there being simple positive properties, and we have good reason to believe that these go no deeper than language. So that leaves us with Imaginability accounts. Imaginability accounts fall into three groups: Liberal, Moderate, and Conservative. But Liberal and Moderate accounts are all superficial, and we've seen that superficial accounts are plagued by a host of problems that deflate their epistemic status: Undecidable Armchair Claims, Concept/Nature Gap, Deep Structure Doubts, Parity, Unexcluded Alternatives, and Dubious Basis. So there is a *prima facie* case for rejecting any account that lets superficial imaginings play the role of justifiers for modal claims. This amounts to a need for a requirement of explanatory detail -- i.e., No Justification Without Explanation. So Liberal and Moderate Imaginability accounts are ruled out. That leaves us with Conservative accounts. Unfortunately, Conservative accounts suffer from the Vacuity criticism, due to our lack of explanatory detail.

Van Inwagen's criticisms of the various accounts of modal epistemology lead to a formidable dilemma for armchair methods of establishing modal claims, stemming from their superficiality: one can either stick with an account that allows for superficiality, or tighten up the requirements to rule it out. If one goes with an account of the former sort, then all the problems listed above arise and sink the account. But if one goes with an account of the latter sort, then although one has thereby ruled out all of those problems, the account is now vacuously true. So we're left with two unsavory alternatives to modal

238

skepticism: implausible superficiality-permitting accounts or plausible but vacuous superficiality-prohibiting accounts. So, pending some other method of modal justification, there is pressure to adopt some form of modal skepticism. Now there are two main forms of modal skepticism: Radical Modal Skepticism and Remoteness Modal Skepticism. But Radical Modal Skepticism goes too far, as it's clear that we have at least *some* modal knowledge – e.g., possibility claims inferred from actuality, "analytic" knowledge of necessities and impossibilities, basic knowledge of humdrum modal claims, and knowledge gotten from a mix of basic modal knowledge and empirical knowledge of how the world is put together. Thus, there are good reasons for preferring Remoteness to Radical Modal Skepticism.

      Let's put it this way. Let's imagine that there is a god who made human beings, and that in designing the part of our cognitive apparatus relevant to the extent of our modal insight, he had the option of choosing from among four possible "settings": High, Middle, Low, and Off. Unfortunately, he didn't tell us *which* setting he selected, and so we're left to figure it out for ourselves. However, he did tell us that he chose from among the four settings, and a rough idea of the extent of modal insight for each setting. Thus, on the "High" setting, the sky's the limit when it comes to justified modal beliefs. We have insight about the possibility of disembodied souls, Anselmian beings, conscious states realized by the population of China, primary colors in addition to the ones we know about, etc. On the "Middle" setting, the sky's not quite the limit when it comes to modal insight, but we're still capable of a decent stock of justified belief about unactualized possibilities – even about many are moderately remote from ordinary experience. We

239

have insight into the possibility of naturally purple cows, translucent sheets of iron, Goldbach's Conjecture, and so on. By contrast, if our capacity for modal insight is at the "Low" setting, then our modal knowledge is pretty unremarkable -- we have some modal knowledge, but it's not very impressive. We have modal knowledge based on semantical, logical, and mathematical rules – "analytic" modal knowledge – and we have inferences from actuality to possibility. We also know humdrum modal claims related the ordinary practical concerns of daily life. We can also extend our modal knowledge a bit by combining our scientific knowledge of how the world is put together with the sorts of modal claims mentioned above. But that's about it -- anything even moderately remote from ordinary experience is unjustifiable. Finally, the "Off" setting means that we have absolutely no modal knowledge whatsoever. We may *think* we do, but we don't. This is Radical Modal Skepticism, on a par with Radical Skepticism about perceptual knowledge.

With this picture before us, we can put Van Inwagen's big picture argument as follows. Many philosophers think that our capacity for modal insight is really set at High for human beings. However, when you examine things a bit more closely, you see that many of our modal tests admit lots of false positives, or underdetermine which modal claim they support, or imply that modal insight has a dubious, or at least troublingly mysterious, basis not shared by our other cognitive faculties, or a combination of these problems. So there are indications that our modal switch must be at a lower setting, and there's a burden of proof on the one who wants to put it higher. How low could it be set? Well, based on the sorts of worries just mentioned, we have reason to think it's set at

240

Low. Why not "Off"? Because it's clear that we do have at least *some* modal knowledge. For example, everything actual is possible; there's a bunch of modal knowledge right there, and it's logically deduced from a solid source, viz., observation. Further, we know a bunch of "analytic" modal truths: those derived from the logic, semantical rules, and the meanings of words, plus truths got via mathematical reasoning. Also, we have some basic, humdrum, non-"analytic" modal knowledge, such as that liquid wine bottles are impossible, that my table could be moved three feet to the left, etc. Finally, we can extend our "basic" modal knowledge by combining it with some basic facts about how the world is put together and reasoning to new modal knowledge. For example, I can combine my basic modal knowledge that things are made out of their "stuff" necessarily with my knowledge that water is H20 to deduce that water is necessarily H20. For these reasons, we have good reason evidence that our "modal setting" is set at Low.

Is there any other evidence that might show that our modal setting is higher -- at Middle or High? We certainly can't rule it out for certain, but there is pressure to provide a principled basis for thinking so. As it turns out, the most rigorous and plausible accounts of justified modal inference entail that the bar can't be set at High, for the accounts require that a modal claim appears or seems to be true at some imagined possible world. But the only way for a modal claim to seem true at a world is for one to have lots of explanatory detail. But no such detail is available in most cases, at least not from the armchair. So, the bar can't be set higher than Middle. So there is a challenge: either give a plausible account of justified modal inference that shows our setting to be at

241

High, or at least Middle, or show that we can get an interesting amount of philosophical work done even if the bar is set at Low. But as our investigation in this dissertation has shown, the prospects for meeting this challenge don't look promising.

242

Appendix III:
Chalmers' Two-Dimensionalist Modal Epistemology: Some Details

1. Concepts

1.1 Primary vs. secondary intensions

According to Chalmers' two-dimensionalism[182] (henceforth *2D*), each concept has two

components: a primary intension and a secondary intension.[183] The *primary intension* of

a concept is a function from scenarios to extensions.[184] Very roughly, a *scenario* is a

comprehensive, consistent hypothesis of a way the *actual* world could turn out to be

---

[182] The following sketch of Chalmers' two-dimensionalist account of modal epistemology is based primarily on two of Chalmers' writings: *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press, 1997), especially ch. 2, and "Does Conceivability Entail Possibility?", in Gendler, Tamar Szabo and John Hawthorne (eds.), *Conceivability and Possibility* (Oxford: Oxford University Press, 2002), pp. 145-200. However, I was also helped by a number of Chalmers' other papers, all but the first of which can be found at his web page, <http://consc.net/chalmers/>: "Epistemic Two-Dimensional Semantics", *Philosophical Studies* 118 (Mar., 2004), pp. 153-226; "The Components of Content", <http://consc.net/papers/content.pdf>; "The Nature of Epistemic Space", <http://consc.net/papers/espace.html>; "The Foundations of Two-Dimensional Semantics", <http://consc.net/papers/foundations.html>; "On Sense and Intension", <http://consc.net/papers/intension.pdf>; and "Two Dimensional Semantics", <http://consc.net/papers/twodim.pdf>.

[183] *The Conscious Mind*, p. 57.

[184] *Ibid.* Strictly speaking, Chalmers here defines the primary intension component as a function from *worlds* to extensions. However, in later writings, he seems to prefer an account that characterizes such intensions in "epistemic" terms, viz., *descriptions* of worlds of a certain sort. See the next footnote for more detail.

243

that's not ruled out by what you know *a priori*.[185] Thus, consider the hypothesis H that

the actual world is one in which everything is the way I believe it to be, except that the

stuff that plays the watery role is not H20, but rather XYZ. I can't rule it out from what I

know *a priori* that H accurately describes the world I live in; H is therefore a scenario.

Now with the notion of a scenario in mind, consider the concept, *water*. The primary

intension of *water* is a function from scenarios to extensions, and thus picks out the class

of things that, for all I know *a priori*, can play the watery role (e.g., H20, XYZ, etc.).

By contrast, the *secondary intension* of a concept is a function from worlds to

extensions *relative to a world held fixed*.[186] Thus, consider *water* again. The secondary

intension of *water* is a function from worlds to the set of tokens that play the watery role

relative to a world held fixed. So if the *actual* world is held fixed, then the secondary

intension of *water* yields the set of H20 tokens; if instead an XYZ-world is held fixed,

then the secondary intension of *water* yields the set of XYZ tokens.

In light of the preceding, we see that primary intensions are more coarse-grained

than secondary intensions. Primary intensions track all epistemically possible occupants

---

[185] In different places, Chalmers has alternately construed scenarios in "metaphysical" terms – viz., in terms of centered possible worlds – and in "epistemic" terms – viz., (roughly) in terms of maximal consistent sets of propositions describing such worlds in an idealized language. He seems to prefer the epistemic construal of scenarios in his more recent writings. See, for example, Chalmers, "Epistemic Two-Dimensional Semantics", especially p.176ff. In any case, since we're primarily interested in Chalmers' modal epistemology, and not his account of semantics, it makes more sense to focus on the epistemic rather than the metaphysical construal of scenarios. For a metaphysical construal would seem to blatantly build reliable connections between meaning and possibility into our language, and thus into his account of conceivability. For these reasons, I give Chalmers' account of scenarios here the epistemic construal.

[186] *The Conscious Mind*, p. 57.

244

of a given role, without distinction. But if so, then if it turns out that there is more than one possible occupant of a given role, then the primary intension is liable to track different referents at different worlds[187]. For they only capture what I shall call *role information* – the primary intension of *being F* means, roughly, 'whatever plays the F-role'. By contrast, secondary intensions capture a specific, perhaps smaller, subset of the possible occupants of a given role. And they can do this in a way that rigidly tracks the same individuals across different worlds. They thus capture what I shall call *occupant information* – the secondary intension of *being F* means, roughly, 'the stuff that plays the F-role *in W*', where 'W' denotes a particular world-held-fixed.

1.2 Comparison: Kaplan, character, context and content

It's helpful to think of Chalmers' 2D framework in terms of an unrestricted version of David Kaplan's treatment of indexicals[188]: *all* expressions – not just indexicals – are amenable to the character/context treatment. For one can use Kaplan's basic two-component approach to zero in on just about any referents one likes by means of two semantic components: a primary intension – which functions, roughly, as a Kaplan-style *character* in the 2D framework – and a world – which functions, roughly, as a Kaplan-style *context* in Chalmers' 2D framework. Together, a primary intension and a world-held-fixed constitute a secondary intension, and thus capture a determinate referent, just

---

[187] And if the space of epistemic possibilities outstrips the space of metaphysical possibilities, the primary intension may have *no* metaphysically possible referent. I will discuss the general issue of Chalmers' version of epistemic possibility as a guide to metaphysical possibility later on, when I get around to evaluating Chalmers' account.

[188] As Chalmers does in (e.g.) *The Conscious Mind*, p. 58-64.

as a character and a context together capture the content of an indexical, and thus capture a determinate referent.

2. Statements

2.1 Primary vs. Secondary Intensions

With Chalmers' account of concepts in mind, we can now explain his account of the meaning of statements. Thus, statements are associated with a primary intension and a secondary intension, where these are composed of the sorts of concepts explained above.[189] Chalmers calls the primary intension of a statement its *primary proposition*, and he calls the secondary intension of a statement its *secondary proposition*.[190] The primary proposition of a sentence is composed of the primary intensions of the concepts associated with that sentence. It is a function from scenarios to truth-values. By contrast, the secondary proposition of a sentence is composed of the secondary intensions of the concepts associated with that sentence. It is a function from worlds to truth-values, relative to a given world held fixed.

To illustrate the distinction between primary and secondary propositions, consider the following sentence:

S. The Morning Star is the Evening Star.

The primary proposition of S, S_{PP}, captures the information of the primary intensions of 'The Morning Star and 'The Evening Star' – i.e., information about the role played by the

---

[189] See, for example, *ibid.*, pp. 63-64, and "Two-Dimensional Semantics", especially section 3.1, "The Core Claims of Two-Dimensionalism".

[190] *The Conscious Mind*, pp. 63-64.

246

Morning Star and about the role played by the Evening Star in all scenarios. A rough approximation of S_{pp} is thus:

S_{pp}. The heavenly body visible in yonder region of the morning sky is the heavenly body visible in yonder region of the evening sky.

Thus, S_{PP} is true in just those scenarios where the occupants of the two roles are identical. But since there are scenarios in which distinct occupants play the roles, S_{PP} is a contingent proposition.

By contrast, the secondary proposition of S, S_{SP}, captures the information of the secondary intensions of 'The Morning Star' and 'The Evening Star'. Thus, if we hold the actual world fixed, it asserts that the thing that plays the Morning Star role in the actual world is the thing that plays the Evening Star role in the actual world; in effect, then, it asserts that

$S_{SP.}$ Venus is Venus.

Thus, S_{SP} is a necessary proposition (given the necessity of identity).

3. Evaluating Intensions

The truth-conditions for primary propositions differ from those of secondary propositions. Accordingly, they differ in their evaluation-conditions. To evaluate the truth-value of a <u>primary</u> proposition at a world, you consider that world as *actual*; to evaluate the truth-value of a <u>secondary</u> proposition at a world, you consider that world as

247

counterfactual.[191] It will be helpful to explain the distinction by employing an example. Thus consider the following sentence:

S. Water is XYZ.

Now consider an XYZ-world, W_{XYZ}. To consider W_{XYZ} as *actual*, I suppose that in the actual world, it turns out that the scientists are wrong, and that the clear, potable liquid that falls from clouds, etc., is really composed of XYZ, and not H20. In other words, I consider a *scenario* of the sort described above. Doing so enables me to evaluate the following indicative conditional:

IC. If W_{XYZ} is actual, then water is XYZ.

Now the primary intension of S is *true* at W_{XYZ} iff IC is true. In this way, considering W_{XYZ} as actual can be used as a method of evaluating S's primary intension.

On the other hand, if I consider W_{XYZ} as *counterfactual*, I hold the actual world, W_{H20}, fixed, and then consider W_{XYZ} as counterfactual to W_{H20}. Doing so puts me in a position to evaluate the following subjunctive conditional:

SC. If W_{XYZ} were actual, then water would have been XYZ.

Now the secondary intension of S is *true* at W_{XYZ} iff SC is true. But since, on our assumption, the reference of 'water' has been fixed at W_{H20}, 'water' rigidly refers to H20. Since this is so, the stuff with the water-stereotype at W_{XYZ} – i.e., XYZ -- is not water, and thus SC is false. In this way, considering a world as counterfactual can be used as a method of evaluating S's secondary intension.

---

[191] On this distinction, see, for example, "Does Conceivability Entail Possibility?", p. 162-164.

248

At this point, we've sketched enough of Chalmers' 2D framework to put us in a position to understand his account of the epistemology of possibility. We will therefore move on to discuss the basic contours of the latter.

4. 2D, conceivability, and possibility

4.1 1-possibility vs. 2-possibility

Chalmers uses his 2D framework to distinguish between two sorts of possibility-statements.[192] Thus, a statement S is *1-possible* iff its <u>primary</u> proposition is true at some possible world, i.e., if it's true at some world considered-as-actual.[193] These are statements about what's metaphysically possible irrespective of what occurs at a world held fixed. Thus, the statement that, possibly, there is XYZ-stuff that plays the watery role is a 1-possibility statement.

By contrast, a statement S is *2-possible* iff its <u>secondary</u> proposition is true at some possible world, i.e., if it's true at some world considered-as-counterfactual.[194] These

---

[192] I should emphasize though, Chalmers' thesis that there is just one type of possibility relevant to both sorts of statements, viz., *metaphysical* possibility. For a primary aim of Chalmers' 2D account is to show that the *a posteriori* necessity cases raised by Kripke and Putnam do *not* establish two types of possibility and necessity: metaphysical and epistemic. For on Chalmers' account, we can make sense of the Kripke/Putnam *a posteriori* necessity cases in terms of just *one* set of worlds, viz., the metaphysically possible worlds. And if we can do that, then philosophers need not abandon the modal rationalism characteristic of traditional armchair philosophy. See Chalmers remarks on this in, e.g., "Does Conceivability Entail Possibility?", p. 162, and pp. 194-95.

[193] The following discussion of possibility is based on Chalmers' remarks in *The Conscious Mind*, p. 67.

[194] *Ibid.*

are the more familiar sorts of possibility statements, viz., those about counterfactual possibilities. Thus, unlike 1-possibility statements, 2-possibility statements are those about what's possible relative to a world held fixed. So, for example, the statement that, possibly, there is XYZ water is a (presumably false) 2-possibility statement.

4.2 Primary vs. secondary conceivability

Relatedly, Chalmers also uses his 2D framework as a basis for distinguishing between two sorts of conceivability. Thus, Chalmers calls a statement *S primarily conceivable* if it's conceivable that *S* is *actually* the case.[195] This is the sort of conceivability involved in evaluating primary intensions via considering a world as actual. Thus, we can use the evaluation-conditions for primary intensions when engaging in primary conceiving: we imagine a world considered as actual, and then see if it verifies the relevant indicative conditional of the form, "If *W* is the case, then *S* is the case." If it does, then one is *prima facie* justified in believing that *S*'s primary proposition is possibly true.

By contrast, Chalmers calls a statement *S secondarily conceivable* if it is conceivable that *S* might have been the case.[196] This is the sort of conceivability involved in evaluating secondary intensions via considering a world as counterfactual. Thus, we can use the evaluation-conditions for secondary intensions when engaging in secondary conceiving: we objectually imagine a scenario considered as counterfactual, and then see if it verifies the relevant subjunctive conditional of the form, "If W were the case, then S

---

[195] Chalmers. "Does Conceivability Entail Possibility?", p. 157.

[196] *Ibid.*

would have been the case." If it does, then one is *prima facie* justified in believing that

S's secondary proposition is possibly true.

4.3 The resultant modal epistemology: a first pass

With Chalmers' distinctions between primary and secondary conceivability, and between

1- and 2-possibility, we're in a position to grasp the fundamental contours of his 2D

account of the epistemology of possibility. Thus, on Chalmers' account, primary

conceivability provides *prima facie* justification for claims about 1-possibility, and

secondary conceivability provides *prima facie* justification for claims about 2-

possibility.[197] However, primary conceivability is a dubious guide to 2-possibility. For it

is liable to lead to modal error, due to the *a posteriori* necessities pointed out by Kripke

and Putnam.

4.4 Positive vs. negative conceivability

So far, we've seen that Chalmers distinguishes between two sorts of possibility claims

and the two corresponding methods of evaluating those claims. We've also seen that

evaluating these differing sorts of claims involves conceiving scenarios of certain sorts.

However, one is no doubt concerned about what conceiving such scenarios amounts to.

For as we've seen, one's account of conceiving can make or break one's modal

epistemology.

Of course, Chalmers is well aware of this concern, and finds promise in the basic

contours of Yablo's imaginability account of conceiving. He thus advances a version of

imaginability that's similar to Yablo's in at least two key respects. First, he accepts

---

[197] On this point, see, e.g., *The Conscious Mind*, p. 67, and "Does Conceivability Entail
Possibility?", p. 165.

251

Yablo's distinction between propositional and objectual imagining, and requires the latter as epistemic backing for the former. Thus, he labels the mere failure to find incoherence within a statement, or its implications, *negative conceivability*, and the act of objectually imagining a scenario that appears to *make* the target modal claim true *positive conceivability*. And when it comes to *justifying* possibility-claims, positive conceivability is the sort of conceivability that's relevant. [198]

Second, he accepts the key elements of Yablo's notion of *verifying* a possibility claim. Thus, he requires that a primary or secondary intension of a statement S be backed not only by an objectual imagining, but also that such an imagining verify the relevant conditional -- indicative for primary intensions; subjunctive for secondary intensions -- if it is to be *prima facie* justified for one as possibly true.[199]

4.5 Prima face vs. ideal conceivability

However, Chalmers' version of imaginability is a bit more nuanced than that of Yablo's. For beyond his two-dimensionalist innovations of the primary/secondary conceivability distinction and the 1-possibility/2-possibility distinction, he makes the further distinction between *prima facie* and *ideal* conceivability. Thus, P is *prima facie conceivable* if an initial fleshing out of an objectually imagined scenario seems to verify P. And P is *ideally conceivable* if, roughly, (i) it is *prima facie* conceivable, and (ii) no amount of

---

[198] Chalmers, "Does Conceivability Entail Possibility?", pp. 149-156

[199] Chalmers, *The Conscious Mind*, p. 67; Chalmers, "Does Conceivability Entail Possibility?", p. 150 ff. Chalmers mentions his indebtedness to Yablo's account here on p. 150.

further fleshing out of the scenario would undercut this seeming verification of P.[200] He

gives an account of the latter notion because he thinks a good case can be made for the

thesis that conceivability *entails* possibility, given his 2D account.[201] Thus, he thinks

other contemporary philosophers who have worked in the sub-field of modal

epistemology, such as James Van Cleve and Stephen Yablo, have construed the link

between conceivability and possibility a bit too weakly, viz., as a relation of fallible

prima facie justification, as opposed to one of entailment. For (thinks Chalmers) they

worry too much about cases of misdescribing worlds, such as occurs in, most saliently,

the *a posteriori* necessity cases.[202] However, given his careful separation of primary and

secondary conceivability, of 1- and 2-possibility, and given his prescription against using

primary conceivability as a guide to 2-possibility, he thinks he can avoid such cases.

---

[200] "Does Conceivability Entail Possibility?", pp. 147-149.

[201] *Ibid.*, p. 171.

[202] *Ibid.*, p. 162; *The Conscious Mind*, p. 367, fn. 32.