

UCLA

UCLA Electronic Theses and Dissertations

Title

Mind, Modality, and Meaning: Toward a Rationalist Physicalism

Permalink

<https://escholarship.org/uc/item/2bd555pm>

Author

Rabin, Gabriel Oak

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

**Mind, Modality, and Meaning:
Toward a Rationalist Physicalism**

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Philosophy

by

Gabriel Oak Rabin

2013

© Copyright by
Gabriel Oak Rabin
2013

ABSTRACT OF THE DISSERTATION

Mind, Modality, and Meaning: Toward a Rationalist Physicalism

by

Gabriel Oak Rabin

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2013

Professor Tyler Burge, Chair

This dissertation contains four independent essays (chapters) addressing a cluster of related topics in the philosophy of mind. Chapter 1: “Fundamentality Physicalism” argues that physicalism can usefully be conceived of as a thesis about fundamentality. The chapter explores a variety of other potential formulations of physicalism (particularly modal formulations), contrasts fundamentality physicalism with these theses, and offers reasons to prefer fundamentality physicalism over these rivals.

Chapter 2: “Modal Rationalism and the Demonstrative Reply to the Master Argument Against Physicalism” introduces the *Master Argument Against Physicalism* and investigates its crucial premise: the inference from an a priori gap between the physical and consciousness to a lack of necessitation between the two.

I argue against the strong form of modal rationalism that underwrites the master argument and offer a more moderate rationalist view. I offer a novel *demonstrative reply* to the master argument, according to which a connection between conscious experience and demonstratives, not dualism, is the source of the epistemic gap between consciousness and the physical.

Chapter 3: “Conceptual Mastery and the Knowledge Argument” argues that Frank Jackson’s famous anti-physicalist knowledge argument featuring Mary, a brilliant neuroscientist raised in a black and white room, founders on a dilemma. Either (i) Mary cannot know the relevant experiential truths because of trivial obstacles that have no bearing on the truth of physicalism or (ii) once the obstacles have been removed, Mary can know the relevant truths.

Chapter 4: “Toward a Theory of Conceptual Mastery” investigates the question “Under what conditions does a thinker fully understand, or have mastery of, a concept?” I argue against three views of conceptual mastery, according to which conceptual mastery is a matter of holding certain beliefs, being disposed to make certain inferences, or having certain intuitions. I propose and respond to objections to my own “meaning postulate view” of the conditions under which a thinker has mastery of a concept.

The dissertation of Gabriel Oak Rabin is approved.

David Chalmers

Mark Greenberg

Tyler Burge, Committee Chair

University of California, Los Angeles

2013

with love and gratitude,
for my parents,
Pia and Marvin

Contents

Introduction	1
0.1 Four Essays	1
0.2 Chapter 1: Fundamentality Physicalism	3
0.3 Chapter 2: Modal Rationalism and the Demonstrative Reply to the Master Argument Against Physicalism	4
0.4 Chapter 3: Conceptual Mastery and the Knowledge Argument . . .	6
0.5 Chapter 4: Toward a Theory of Conceptual Mastery	8
1 Fundamentality Physicalism	10
1.1 Fundamentality	11
1.1.1 The Building Blocks of Reality	11
1.1.2 Two kinds of creation	14
1.1.3 The Relata of Metaphysical Generation	16
1.1.4 The Covering Constraint	17
1.1.5 Generation and Ground	19
1.2 Fundamentality Physicalism	23
1.2.1 Fundamentality Physicalism	23
1.2.2 Fundamentality Physicalism and the Mathematical	25
1.3 Modal Physicalism	28

1.3.1	Two Criticisms of Modal Formulations of Physicalism	34
1.3.2	Fundamentality Physicalism and Modal Physicalism	35
1.4	Other Physicalisms	39
1.4.1	Everything is Narrowly or Broadly Physical: A Schema	39
1.4.2	Applying the Schema to Fundamentality Physicalism	40
1.4.3	Applying the Schema to Modal Physicalism	41
1.4.4	Type Identity Physicalism	42
1.4.5	Realization Physicalism	43
1.4.6	Nature Physicalism	45
1.4.7	Explanatory Physicalism	46
1.4.8	Definition Physicalism	48
1.4.9	Concept Physicalism	49
1.4.10	Summary	52
1.5	Fundamentality Physicalism and Other Physicalisms	53
1.5.1	Against Nature and Explanatory Physicalism	53
1.5.2	The Relevance of Fundamentally Non-Physical Worlds	54
1.5.3	Fundamentally Non-Physical Worlds are Metaphysically Pos- sible	55
1.5.4	Space and Time	57
1.5.5	Lessons from the Possibility of Fundamentally Non-Physical Worlds	58
1.5.6	Fundamentality Physicalism and the Fundamentally Non- Physical	61
1.6	Conclusion	62

2	Modal Rationalism and the Demonstrative Reply to the Master Argument Against Physicalism	64
2.1	Introduction	65
2.2	The Master Argument Against Physicalism	67
2.2.1	From Fundamentality to Modality to A Priori Implication	67
2.2.2	The Master Argument	71
2.2.3	Premise (P3): From Fundamentality to Modality	72
2.2.4	Premise (P1): The Epistemic Gap	72
2.2.5	Premise (P2): From an epistemic to a metaphysical gap	75
2.3	Supplementation	76
2.3.1	Covering Your Bases	76
2.3.2	The Interest of the Base	79
2.3.3	The Supplementation Question	82
2.3.4	Naive Modal Rationalism	83
2.3.5	Strong Modal Rationalism	85
2.4	Return to the Master Argument	87
2.4.1	The Strong Modal Rationalist’s Test for Fundamentality	87
2.4.2	Canonical Expression of the Base: Semantic Neutrality	89
2.4.3	Semantic Neutrality and the Master Argument	95
2.5	Motivations for Modal Rationalism	96
2.5.1	Overview	96
2.5.2	Some Ideas Behind Modal Rationalism	97
2.5.3	The View From Above	100
2.5.4	The Strong Modal Rationalist Link and the View From Above	102
2.6	Against Strong Modal Rationalism	102

2.6.1	The Recipe View	102
2.6.2	The Trench View	105
2.6.3	The Recipe View and the Strong Modal Rationalist Link	111
2.7	The Demonstrative Reply to the Master Argument Against Physicalism	116
2.7.1	The Demonstrative Reply	117
2.7.2	Comparison to other Demonstrative Oriented Replies	121
2.7.3	The Ineliminable Connection Between Phenomenal Concepts and Demonstratives	126
2.7.4	Phenomenal Concepts, Demonstration, and the Experiential Requirement	132
2.7.5	Toward A More Moderate Rationalism: the Moderate Rationalist Link	133
2.7.6	Objection: The Hard Problem of Cats	136
2.8	Conclusion	137
3	Conceptual Mastery and the Knowledge Argument	139
3.1	Introduction	140
3.2	Physicalism	140
3.3	The Knowledge Argument	142
3.4	Roadmap	145
3.5	The Missing Concept Reply	148
3.6	The Conceptual Mastery Reply	151
3.7	Aside: The Phenomenal Concept Strategy	155
3.8	The Informational Assumption and the Second Horn	160
3.9	Lonely Mary and the Pincer Argument	164

3.10	Objections	171
3.11	Taking Stock	175
3.12	Conclusion	179
4	Toward a Theory of Conceptual Mastery	181
4.1	Introduction	182
4.2	Preliminaries	183
4.2.1	Terminology and assumptions	183
4.2.2	Cores and Conceptual Roles	186
4.2.3	Precisifying the Notion of Conceptual Mastery	188
4.2.4	Against Skepticism	190
4.2.5	Four Theories of Conceptual Mastery	194
4.3	Deviant Masters: Against the Necessity of Belief, In- tuition for Conceptual Mastery	197
4.3.1	Overview	197
4.3.2	The Argument	197
4.3.3	Response 1: Agnes remains disposed to infer	199
4.3.4	Response 2: BOCHE is not a concept	200
4.3.5	A Counterexample Without a Defective Concept	201
4.3.6	Extending the argument against the belief and intuition views	202
4.4	The Special Status Problem: Against the Sufficiency of Belief, In- ference, or Intuition for Conceptual Mastery	202
4.4.1	Overview of the Special Status Problem	202
4.4.2	Not Merely Belief	203
4.4.3	Sophisticated Views	204
4.4.4	The All and Only Problem	205

4.4.5	Moving On	207
4.5	The Meaning Postulate View	208
4.5.1	From the Failures of Rivals to the Meaning Postulate View	208
4.5.2	Taking to Govern Use	209
4.5.3	Governing Use	212
4.6	Three Objections	214
4.6.1	Overview of the Objections	214
4.6.2	Objection 1: Against the notion of “taking a rule to govern the use”	214
4.6.3	Objection 2: Mastery Without Taking to Govern Use	216
4.6.4	Objection 3: Taking to Govern Use Without Mastery	217
4.7	Conclusion	220

Bibliography		221
---------------------	--	------------

Acknowledgements

This dissertation would not be possible without assistance and support from a wide variety of individuals over the years. First and foremost, I would like to thank my parents, Pia and Marvin, to whom this thesis is dedicated. They never pushed me to become a lawyer or a doctor or to pursue a career more respectable by traditional standards. They supported me in whatever wild endeavour I chose to pursue. When I skipped about from anthropologist to mathematician to wilderness adventurer to poet to ski bum, their support did not waver (even if their eyebrows did). The same held true when I opted for graduate school in philosophy. Thanks Mom and Dad.

Other members of my family played important roles in my success as well. My brothers Micah and Noah are two of the best friends and comrades one could ask for. Whatever the occasion required, they were always ready with a sympathetic ear, a funny story, or a cold beer. My aunts, uncles, and cousins deserve gratitude as well. I'd particularly like to thank my grandparents Jack and Millie. Their support - both personal and financial - enabled a poor kid from Vermont with big dreams to pursue whatever intellectual pipe dream caught his fancy.

On the academic side, the members of my dissertation committee offered much needed guidance and support throughout the process of writing this dissertation and my graduate education more generally. Tyler Burge, both in person and through his writing, imparted much philosophical wisdom. In many long

meetings, he provided philosophical stimulation and direction, a stern hand and strategic advice, and a healthy no nonsense philosophical attitude that kept me grounded. Sometimes I bristled at his demands, but the finished product is that much better for them. His extensive comments on many drafts helped polish and shape this dissertation in a multitude of ways.

David Chalmers deserves credit for introducing me first through his writing to the major issues this dissertation addresses, for making seminal contributions to those topics, and later for serving as an invaluable philosophical advisor and interlocutor. Much of this dissertation targets David's work. But my criticisms demonstrate my respect and admiration for his contributions to philosophy. David and I agree on many things, and I think of myself in as defending David's big picture approach to philosophy and specifically the mind-body problem. In many ways, this dissertation, and my approach to the philosophical issues in it, especially the nature of mind and of representational content, are the bastard child of the Tyler Burge and the David Chalmers approaches. They are both intellectual giants. I am proud to count myself among their philosophical progeny.

I thank Mark Greenberg for being an invaluable philosophical interlocutor. Mark and I think and operate philosophically in similar ways, and it was always helpful for me to bounce ideas off him before, during, and after, any philosophical project. Mark also deserves thanks for useful comments on a variety of papers, for his mentorship, and for teaching me many important lessons in meta-semantics and the philosophy of representation more generally. I always walked away from his stimulating seminars amped up to write something new. The last essay in the dissertation - "Toward a Theory of Conceptual Mastery" - in partic-

ular owes a great deal to his tutelage.

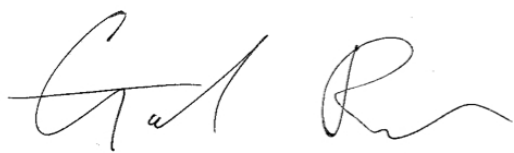
Calvin Normore is a philosophical master who seems to know about every topic in philosophy. He was an invaluable resource, always ready with a comparison between whatever problem interested me and an issue addressed by Antoine Arnauld, William of Ockam, or John Buridan. He picks up material quickly and always offers some interesting angle on the problem. I owe him thanks for getting me to consider issues from a different angle and helping me move past logjams in my thinking.

A philosopher's development owes as much to the philosophical environment in which he or she finds him or her self as to the particular instructors, tutors, and courses he or she encounters. I want to thank all the classmates and colleagues I've had over the years, at UCLA, UMass Amherst, ANU, NYU, and Columbia. Brian Rabern, Denis Buehler, David Friedell, Rachel Johnson, Ole Koksvik, Kelvin McQueen, Eliot Michaelson, David Friedell, and Tamar Weber provided invaluable friendship and stimulating philosophical interaction. Einar Bohn, Emily Curd, Jeremy Cushing, Billy Dunaway, Jeffrey Dunn, Edward Elliot, Edward Ferrier, Mikkel Gerken, Ebru Kardan, Brandt van der Gaast, Jonathan Simon, and Clas Weber deserve thanks as well. Amongst faculty, I'd like to thank Joseph Almog, Torin Alter, Ned Block, Paul Boghossian, Phillip Bricker, Bill Brewer, Sam Cumming, Kit Fine, Gabe Greenberg, Pamela Hieronymi, David Kaplan, Kevin Klement, Hilary Kornblith, Stephan Leuenberger, David Manley, Terry Parsons, Jonathan Schaffer, Ted Sider, Sheldon Smith, Daniel Stoljar, and Wolfgang Schwarz. I want to thank David Albert, Bernard Berofsky, John Collins, and Akeel Bilgrami for getting me into philosophy in the first place during my days as a mathematics undergraduate at Columbia. John Conley, Seth

Gale, Melissa Donner, and Ilan Wapinski were valuable friends and stimulating philosophical conversation mates during those early days at Columbia when I was finding my path. Finally, I'm sure to have forgotten someone. For that I apologize and attach a "thank you".

Thanks to everyone who has helped and aided in the philosophical and personal progress that has culminated in this dissertation. It remains a work in progress that has, for now, reached a resting point. The faults and errors contained herein are my own, and appear in spite, and not because, of the influence of the aforementioned persons.

Sincerely,

A handwritten signature in black ink, appearing to read "Gabriel Rabin". The signature is written in a cursive style with a large initial "G" and a long, sweeping tail on the "R".

-Gabriel Rabin

Vita

Education

- C.Phil Philosophy, University of California, Los Angeles, 2010
- M.A. Philosophy, University of California, Los Angeles, 2008
- University of Massachusetts, Amherst, 2004-2006 (no degree awarded)
- B.A. Mathematics, Columbia University, New York, 2002

Employment

- Teaching Assistant, University of California, Los Angeles, 2007-2013
- Teaching Assistant, University of Massachusetts, Amherst, 2005-2006
- English and Mathematics Teacher, The Lowell Whiteman School, Steamboat Springs, Colorado, 2002-2004

Publications

- "Conceptual Mastery and the Knowledge Argument" in *Philosophical Studies* 54: 125-147, 2011.
- "Full-Blooded Reference" in *Philosophia Mathematica* 15 (3): 357-365, 2007.
- Review of Daniel Stoljar's "Physicalism" in *Australasian Journal of Philosophy* 89 (3): 562-566, 2011.

Presentations and Comments

Presentations

1. "Conceptual Mastery and the Knowledge Argument", New York University, March 8th, 2013.
2. "Well-founding Grounding Grounding" at the Conference on the Philosophy of Kit Fine, Sinaia, Romania, May 25th-27th, 2012.
3. "Toward a Theory of Conceptual Mastery" at the Albritton Society, UCLA, January 27th, 2012.
4. "Supplementation, Meta-semantics, and Modal Arguments Against Physicalism" at The Australian National University, July 12th, 2011.
5. "Grounding Grounding" at the Australasian Association of Philosophy Annual Conference, Dunedin, New Zealand, July 7th, 2011.
6. "Blocking the Blocking Defense" at the Pacific APA, San Diego, April 22nd, 2011.
7. "Modal Rationalism Strikes Back" and "Not All Facts are Physical Facts" at the Albritton Society, UCLA, January 15th, 2011.
8. "Conceptual Mastery and the Knowledge Argument" at the Washington Square Circle, NYU, December 13th, 2010.
9. "Conceptual Mastery and the Knowledge Argument" at MindGrad 2010, the University of Warwick, December 4th, 2010.
10. "Two-Dimensional Meta-Semantic" at The Australian National University, December 17th, 2009.
11. "Physicalism and the Idealization of A Priori Entailment" at the The Australian National University, August 18th, 2009.

12. "Physicalism and the Idealization of A Priori Entailment" at the Australasian Association of Philosophy Annual Conference, Melbourne, July 7th 2009.

Comments

1. Comments on Renee Jorgensen's "A Prosententialist Account of Vagueness", 6th USC/UCLA Graduate Student Philosophy Conference.
2. Comments on Nora Berenstain's "Necessary Laws and Chemical Kinds: A Case Study in A Posteriori Necessity", 4th USC/UCLA Graduate Student Philosophy Conference.
3. Comments on Elia Zardini's "A Paradox of Higher-Order Vagueness" at the 2nd USC/UCLA Graduate Student Philosophy Conference.

Fellowships, Honors, and Awards

- Distinguished Teaching Award (2012) - for best graduate student instructor in philosophy
- Honorable Mention for Distinguished Teaching Award (2010)
- Dean's Humanities Fellowship (2010)
- Fulbright Fellowship (2008)
- U.S. Alumni Fulbright Scholar (2008) - for top-ranked Fulbright applicant
- Summer Research Mentorship (2008, 2010)
- University Fellowship (2007)
- Chancellor's Prize (2006)
- B.A. *Magna Cum Laude* - Mathematics (May 2002)

Introduction

0.1 Four Essays

This dissertation contains four essays addressing a cluster of related topics in the metaphysics of mind and the nature of representation, particularly conceptual representation. Officially, the essays are independent from each other. Each can be read and evaluated on its own merits. However, the essays complement each other and together form a coherent picture, but not a sequential narrative.

The dissertation takes as its starting impetus a particular problem in the metaphysics of mind. This problem stems from the so-called “hard problem” of consciousness, and was forcefully brought into the contemporary philosophical discussion by David Chalmers’ seminal work *The Conscious Mind* [1996].¹

The problem stems from a triad of claims that are in tension. (The presentation is mine, not Chalmers’).

(EG) Epistemic Gap: There is an epistemic explanatory gap between the physical and conscious experience.

(Phys) Physicalism: Our world is, in some important sense, entirely physical. This entails, at the least, a modal connection between the physical and

¹Chalmers’ work had important precursors in Jackson (1982), Kripke (1972), Levine (1983) and Nagel (1974), among others. But Chalmers put the problem in its most pressing and direct form. His work shaped the dialectic in the years that followed in a manner that the other works did not.

everything else.

(MR) Metaphysical Rationalism: There are important and close connections between metaphysical notions such as necessity, fundamentality, ground, and metaphysical generation on the one hand, and epistemic notions like a priority, conceptual truth, knowability, and explanation, on the other.

The tension is easy to see. Epistemic Gap claims an epistemic explanatory gap between the physical and consciousness. Physicalism denies a metaphysical gap. Metaphysical Rationalism connects epistemic and metaphysical gaps, and licenses an inference from the epistemic explanatory gap between the physical and conscious experience to a metaphysical gap. In order to make the tension clear, one must precisify the nature of the epistemic and metaphysical gaps, and the status of the connection between the two that metaphysical rationalism posits.

The form of metaphysical rationalism most salient in this dissertation is *Modal Rationalism*, according to which there are important, potentially constitutive, connections between metaphysical modality and the a priori. Modal Rationalism underwrites the key premise in *the master argument against physicalism*, which I discuss in chapter 2 (“Modal Rationalism and the Demonstrative Reply to the Master Argument Against Physicalism”).

One can reply to the tension, and to the master argument (which precisifies the tension), in a variety of ways. Of course, one can give up physicalism and opt for dualism (Chalmers (1996a)). One can deny the epistemic gap between the physical and conscious experience, perhaps maintaining that it is closable “in principle” (Braddon-Mitchell (2003), Jackson (2005)). But by far the most popular response is the rejection of metaphysical, or modal, rationalism (Block & Stalnaker (1999), Hill (1997), Levine (1983, 2004), Loar (1990), Papineau (1993),

many others). One, and perhaps the, overarching theme of this dissertation is the search for a resolution of the tension that does not give up metaphysical or modal rationalism. I prefer a rationalistically respectable form of physicalism. Chapters 2 and 3 both offer rationalistically respectable responses to prominent arguments against physicalism (the master and knowledge arguments, respectively) that have often been met with a rejection of rationalism altogether.

0.2 Chapter 1: Fundamentality Physicalism

Chapter 1, “Fundamentality Physicalism”, seeks to clarify the thesis of physicalism (or the many theses that might go under that name). I flesh out my favored conception of physicalism, *fundamentality physicalis*, which is based on the notions of *fundamentality* and *metaphysical generation*, and argue that physicalism so defined is what is at stake in many of the contemporary discussions of physicalism. I contrast and compare fundamentality physicalism with *modal physicalism*, which defines physicalism using modal tools. I argue that fundamentality physicalism is superior to modal physicalism and better captures what many philosophers have had in mind.

Fundamentality physicalism entails modal physicalism, but not vice versa. This makes fundamentality physicalism vulnerable to the various arguments that have been launched against modal physicalism, such as the conceivability argument (Chalmers (1996a, 2010) and the knowledge argument (Jackson (1982, 1986))), and links up fundamentality physicalism to the large body of literature on the metaphysical and particularly modal difficulties for physicalism resulting from the epistemic gap between the physical and consciousness.

0.3 Chapter 2: Modal Rationalism and the Demonstrative Reply to the Master Argument Against Physicalism

Chapter 2, “Modal Rationalism and the Demonstrative Reply to the Master Argument Against Physicalism”, is the longest of the chapters. I lay out the master argument against physicalism, which goes roughly as follows. First premise: There is no a priori route from the physical truths to the truths about conscious experience. Second premise: If there is no a priori route from the physical truths to the truths about conscious experience, there is no necessitation from the physical truths to consciousness truths. Third premise (minor): If there is no necessitation from the physical to consciousness, physicalism is false. Conclusion: Physicalism is false.

I demonstrate how many of the contemporary argument against physicalism, including conceivability arguments based on zombies and inverted qualia (Chalmers (1996a, 2010, 2003a)) the knowledge argument (Jackson (1982, 1986)), the explanatory gap argument (Levine (1983)), and Kripke’s modal argument (Kripke (1972)), can be viewed as variations on or implementations of the master argument. I believe that unifying the arguments in this way sheds light on key issues.

The crucial premise of the master argument is the second, which links the epistemic a priori gap between the physical and consciousness to a metaphysical modal gap between the two. I focus most of my attention there. Modal rationalism underwrites inferences from an a priori gap to a modal gap, of which the second premise is an instance. I discuss at length the motivations behind modal rationalism and the connection between modal rationalism and the master argu-

ment.

I argue that rationalists should prefer a different, and weaker, form of modal rationalism, which I call “weak modal rationalism”. Importantly, weak modal rationalism does not license the second premise of the master argument. In addition, weak modal rationalism offers a physicalistically and rationalistically acceptable story about the source of the epistemic gap between the physical and consciousness.

One of my key ideas is that demonstratives, and particularly demonstrative applications of representational tokens (such as pointing at a lake and stating, “*That* is water”), play a vital role in determining reference. Demonstratives are, in a certain sense, representationally fundamental. In particular, demonstratives are essential for locking our representations onto conscious experience. To describe to someone which type of experience we intend to designate, eventually we must point: “*That’s* the sensation I’m talking about!”

In the end, I suggest that the epistemic gap between the physical and consciousness (including difficulties in finding an a priori route from the physical to consciousness) stems from this special connection between consciousness and demonstratives, rather than in a metaphysical gap indicative of dualism. Returning to a recurring theme, I note that this reply to the master argument preserves, rather than abandons, the rationalist motivations with which I have sympathy. We resolve the tension between modal physicalism, metaphysical/modal rationalism, and the epistemic gap by a well-motivated relaxing of the rationalist connection between modality and the a priori.

0.4 Chapter 3: Conceptual Mastery and the Knowledge Argument

Precisifying Epistemic Gap, Metaphysical Gap, and Metaphysical Rationalism involves fixing an epistemic notion of entailment, a metaphysical notion of entailment, and a rationalist story about the connection between the two. “Conceptual Mastery and the Knowledge Argument” uses “in a position to know” as its epistemic notion and necessitation as its metaphysical notion. The relevant form of metaphysical rationalism claims that if the physical facts necessitate all the facts, then the physical facts put one in a position to know all the facts. But Jackson (1982, 1986)’s famous thought experiment involving Mary, a color-deprived neuroscientist raised in a black and white room, seems to show that knowledge of the complete physical facts does not put one in a position to know all the facts. In particular, it does not put one in a position to know the facts about conscious experience. Mary lacks knowledge of important facts about conscious experience, including what it’s like to see red. By the active form of metaphysical rationalism, this epistemic gap entail that the complete microphysical facts do not necessitate all the facts. Modal physicalism claims (more or less) that a complete microphysical truth does necessitate all the facts. Therefore, modal physicalism is false. This is the knowledge argument against physicalism.

The chapter proceeds by agreeing to play the metaphysical rationalist’s game. We admit that, in general, one can get results about which facts necessitate which other facts from testing what idealized epistemic agents such as Mary can and can’t know on the basis of those facts. However, if we’re going to play this game, we must lay down some basic ground rules. Otherwise we’ll rapidly be led into

trivial errors. I argue that the knowledge argument can't succeed within the confines of these basic ground rules.

The most important "ground rule" is that we should only draw conclusions about necessitation from the cases of epistemic agents that have conceptual mastery of all the relevant concepts. Conceptual mastery is something like full understanding of a concept.² I then argue that the knowledge argument founders on a dilemma. Either (i) Mary cannot know the relevant truths about consciousness because she lacks conceptual mastery of phenomenal concepts, in which case no metaphysical conclusions about necessitation or physicalism can be drawn, or (ii) once Mary does have conceptual mastery, she will be able know the missing truths. (A *phenomenal concept* categorizes experiences according to what it is like to be the subject of that experience.)

Two features of this chapter are worth pointing out. First, there are a variety of available responses to the knowledge argument.³ One common move is to reject the rationalist inference from an epistemic to a metaphysical gap. My solution to the knowledge argument stands out for its compatibility with metaphysical rationalism. Second, this chapter connects to the dissertation chapter titled "Toward a Theory of Conceptual Mastery" (to be discussed shortly). There I offer the beginnings of a theory of the conditions under which a thinker has conceptual mastery of a concept.

²I explore the notion of conceptual mastery more fully in Chapter 4: "Toward a Theory of Conceptual Mastery".

³The ability response appears in Lewis (1988) and Nemirow (1990), the acquaintance response in Conee (1994), and the "old fact new mode" response in Byrne (2002) and Tye (1995). These are not the only available responses. Cf. Stoljar (2005) or Nida-Rumelin (2009) for an overview.

0.5 Chapter 4: Toward a Theory of Conceptual Mastery

The 4th chapter, “Toward a Theory of Conceptual Mastery”, is the black sheep of the four chapters. It does not address directly any of the metaphysical issues stemming from the hard problem of consciousness. “Toward a Theory of Conceptual Mastery” is more of a spin-off, or a further exploration of issues first introduced in, Chapter 2, “Conceptual Mastery and the Knowledge Argument”. In that paper, I introduce the notion of “mastery” or “full understanding” of a concept. This chapter seeks to further explore that notion, and specifically to answer the question, “Under what conditions does a thinker have mastery of her concept?”

Concepts are the basic units of thought. The thought OSTRICHES LIKE CHOCOLATE is composed of the concepts OSTRICHES, LIKE, and CHOCOLATE. An agent *possesses* a concept when he or she can think thoughts of which the concept is a component. An agent has *conceptual mastery* of a concept when he or she fully understands that concept. One can possess a concept without having mastery. Many people use technical concepts that have worked their way into public consciousness without fully understanding them (examples include DARK MATTER, CHAOS THEORY, CLOUD COMPUTING, TRANSISTOR, ID, FRACKING, and HEDGE FUND). Any theory of conceptual mastery must answer the following fundamental question: “Under what conditions does an agent have conceptual mastery?” This is the question I attempt to answer in the 4th chapter.

I argue for what I call “the meaning postulate view” of conceptual mastery, according to which mastery of a concept is a matter of taking certain core rules to govern the use of the concept. These rules operate in a manner similar to meaning

postulates in semantics. The rules contribute to giving the concept its meaning.⁴ I argue against “the belief view”, “the inference view”, and “the intuition view”, according to which mastery of a concept involves believing, inferring, or intuiting in accordance with certain propositions or inferences.

⁴I use ‘meaning’ in the generous sense, to cover any variety of representational quality, not just the representational features of language. On this approach, non-linguistic mental states, such as (some) conceptual thoughts and perceptual states, have meaning.

Chapter 1

Fundamentality Physicalism

Chapter Abstract

This essay has three goals. The first is to introduce the notions of *fundamentality* and *metaphysical generation* and to argue that physicalism can usefully be conceived of as a thesis about fundamentality. Toward this end, I argue (i) that fundamentality physicalism has advantages over currently popular modal formulations of physicalism and (ii) that fundamentality physicalism is what many who endorse modal formulations of physicalism had in mind all along. The second goal is to explore several other potential formulations of physicalism, including as a thesis about identity, realization, natures, explanation, language, and concepts. The third goal is to contrast fundamentality physicalism with these other theses and offer reasons to prefer fundamentality physicalism over these other formulations.

1.1 Fundamentality

1.1.1 The Building Blocks of Reality

The world has more and less fundamental aspects. Quarks, leptons, bosons, and their masses, charges, and spins, are likely among the fundamental aspects. Coffees, cars, and koalas are not. The world's fundamental layer is the bedrock of reality. It *metaphysically generates* the rest of the world's contents. The non-fundamental *metaphysically depends* on the fundamental. The coffees, cars, and koalas are the way they are because of the way the fundamentalia are. With the distinction between the fundamental and the non-fundamental in place, an obvious and intuitive question arises: "What is fundamental?" I find this question one of the most natural questions in all of human inquiry. For the most part, physics (tinged with some philosophy) is in the business of answering it.

Fundamentalia come in several (perhaps many) varieties. At the least, there are objects, properties, relations, and laws. (From here, it should be understood that 'properties' includes relations of any adicity). There may be other varieties of fundamentalia as well, but the types I've mentioned seem like a good start.¹

It was once thought that atoms were fundamental. This has turned out not to be so. Atoms are constructed from combinations of electrons, protons, and neutrons, along with some forces keeping them together. Neither protons nor

¹Two other types of potential fundamentalia include *events* and *boundary conditions*. Certain physical equations, such as the field equations of general relativity, are not well-behaved, in the sense of having unique solutions, in the absence of certain assumptions. These assumptions are called 'boundary conditions'. One can reasonably ask why, and if, these boundary conditions hold. For example, the assumption of an asymptotically flat space-time in general relativity ensures that the field equations of general relativity are well-behaved. One might claim that these boundary conditions, or boundary conditions in other domains, are themselves fundamental features of the universe. Cf. Smeenk (2014) for more on the status of boundary conditions. Thanks to Sheldon Smith here.

neutrons are fundamental - they are composed of still smaller particles. The best science at this moment - the standard model - still suggests that electrons are fundamental, though there are open questions about whether each electron is merely a vibration in a string, or a point assigned property values in a field. The details of these proposals needn't concern us here.

Fundamentalia are the basic building blocks of reality.² We need to separate two questions about the fundamental. The first - *the roster question* - is: "What types of fundamentalia are there?" What types of building blocks are there? The second - *the layout question* - is: "What tokens of fundamentalia are there, and how are they?" How are the building blocks actually arranged? The layout questions asks how many fundamentalia there are, what properties they have, and how they arranged across (for example) space-time. Once we've decided that electrons and their spins are fundamental (thereby partly answering the roster question), there remain layout questions about how many electron there are, where they are, what spins they have, and the like.

I think there is a perfectly good notion of "metaphysical building block". These building blocks metaphysically generate other items. Those items in turn metaphysically depend, for their existence and properties, on the building blocks that generate them. My notion of "metaphysical building block" is something like a generalization of the normal notion of "building block". Consider a lego model of the Space Shuttle *Endeavour*. The model is constructed from the individual legos that form its wings, tail, fuselage, and cockpit. These legos, in turn,

²In this paper, I assume that there are ultimate building blocks that constitute a maximally fundamental layer. One might instead think that there is no bottom level. On such a picture, each metaphysical level of reality has further, more fundamental structure underpinning it. One might call such a scenario "infinite descent". Not much turns on this issue. However, it's easier to speak as if there is a maximally fundamental level, so I'll continue to make the assumption.

are constructed from molecular plastic compounds. The lego spaceship is the way it is in virtue of the way the individual legos are, which are the way they are because of the way the plastics are, which, eventually, are the way they are in virtue of how the sub-atomic particles in the vicinity of the spaceship are. However, the state of the lego spaceship is not merely determined by the pieces, the *objects*, that make it up. It's also determined by *how* those pieces are, i.e. by what properties they have. The sub-atomic particles and their properties generate the molecules, compounds, lego pieces, and eventually the *Endeavour* model itself. The metaphysical generators of the model include both objects and properties.

However, talk of fundamentalia as "building blocks" should be not be taken overly literally. Fundamentalia need not be, in any reasonable sense, *part* of the non-fundamentalia they generate. There is a clear sense in which an electron in the koala's nose is part of the koala. This electron and its properties partly generate the koala. But those particles, in order to make a koala, must have a certain cohesiveness across time. If they momentarily coalesce into a koala shape and then disperse, there is no koala. The koala's existence is partly generated by the cohesiveness of the particles, and this cohesiveness is partly generated by the laws that govern the interaction of those particles. With a different set of laws, that very same koala-shaped arrangement of particles would not be a koala. It might instead be a koala-shaped bomb, about to explode. Thus the law that ensures that the koala-shaped particles remain a cohesive unit also partly generates the koala. The law is one of the metaphysical building blocks of the koala. But there is no reasonable sense in which the law is a *part* of the koala.

One can get a grip on the notions of fundamentality and metaphysical generation by imagining oneself tasked with the job of creating the universe. One must

create all the various fundamental particles, assign those particles fundamental properties, and set out the laws that govern their interaction. In so doing, one also creates the non-fundamentalia: the coffees, cars, and koalas. The creation of the koala is not an additional act. Compare to the creation of the lego model of *Endeavour*. Once you arrange all the legos appropriately, you're done. There's no additional separate task of creating the lego model itself. Let's call the test for fundamentality embodied in this thought experiment *the creation test*. Something is fundamental at world w if and only if it would need to be directly created by a being creating w . A creator can indirectly create koalas by creating appropriately arranged fundamentalia. But the fundamentalia must be created directly, because they are not metaphysically generated by anything else.

The search for the roster of fundamentals is similar to Aristotle's search for the primary substances. In *Metaphysics*, Aristotle writes that "Substance is the subject of our inquiry; for the principles and the causes we are seeking are those of substances. For if the universe is of the nature of a whole, substance is its first part..." (1984: 1688; *Meta*.1069a18-20). Aristotle is clear that the substances are ontologically or metaphysically prior to the non-substances. Gill (1989) summarizes Aristotle's test for substance-hood as follows: "the main criterion [for selecting the primary substances] is ontological priority. An entity is ontologically primary if other things depend for its existence on it, while it does not depend in a comparable way on them." Similarly, the fundamentalia are that on which everything else metaphysically depends, and which depend on nothing further.

1.1.2 Two kinds of creation

The generation of a non-fundamental from the fundamental is importantly different than Kyle's generation of a cake from 2 cups of flour, a half-stick of butter,

two eggs, a cup of sugar, baking powder, 1/2 cup of cocoa, and a teaspoon of vanilla extract. Consider the cake, cooling on the kitchen counter, replete in all its chocolaty glory. One can tell two stories about the cake. The first story is causal. It describes Kyle's toils in the kitchen, his stirring of milk and flour, precise placement on the center oven rack at 175 degrees, and the chemical reactions that occur while it bakes. The second story is metaphysical. This story starts from the electrons and bosons inside the cake, proceeds up through the hydrogen, oxygen, and sodium carbonate molecules, and then finally up to the cake itself. The cake itself, with its chocolatey-ness, shape, and texture, metaphysically depends on these smaller and more fundamental elements. When I speak of metaphysical dependence, and generation, I'm talking about the type of dependence and generation in the second story. Clearly, there is a sense of 'dependence' and 'generation' according to which the cake depends on and is generated by Kyle's work and the chemical reactions of baking. But that's not the sense of generation or dependence I'm trying to latch on to.

The picture is somewhat complicated by the fact that it's not unreasonable to think that the cake metaphysically depends on some items on which it also causally depends. Perhaps the sugar and flour provide examples. One might claim that the *metaphysical* generation I'm after is always synchronic, whereas *causal* generation is always diachronic. But I don't think this is quite right. It seems to me that some present item could metaphysically depend, for being what it is, on past events or items. If one endorses an essentiality of origin thesis for cakes, then the cake metaphysically depends on past events, such as the particular process of cooking that yielded that cake. Thus, I think we should leave open the possibility that there are diachronic relations of metaphysical dependence and

generation. However, the synchronic vs. diachronic distinction nonetheless provides a useful heuristic for distinguishing between causal and metaphysical generation.

1.1.3 The Relata of Metaphysical Generation

What are the relata of metaphysical generation? We might remain agnostic, adopting a strategy similar to Fine (2012)'s approach to ground, a relation similar to metaphysical generation. According to Fine, we can make true statements of the form 'sentence₁, sentence₂, ... grounds sentence₃', without making commitments about what the genuine relata of grounding are, or reifying the grounding relation itself. We need claim only that statements of the above form are true.

However, I'm offering a metaphysical picture and an approach to fundamentality. Such pictures are better fleshed out and more informative when they are more committal, even if these commitments come with warts. At the least, I take there to be relations of metaphysical grounding between something like *states of affairs*. A *state of affairs* is a specific way the world is. States of affairs can include the having of properties by objects or the holding of laws. The state of affairs of some electrons and protons being organized thus and so, with these charges and masses, metaphysically generates the state of the atom's having the mass it does. Some states of affairs are more fundamental than others. The states of affairs that are generated by no further states of affairs are the fundamental states of affairs. We also count as fundamental whatever is involved in those states of affairs, be they objects, properties, events, laws, or what have you.

Metaphysical generation relates a collection of states of affairs to a single state of affairs. (This does not preclude the possibility that the collection contains only one element). The states of affairs, together, generate another state of affairs.

For example, the electrons having such and such mass, the protons having such and such mass, etc., metaphysically generates the atom's having its mass. A given collection can generate many other states of affairs. But it generates each one individually, rather than them all as a collection.

One might also wish to posit relations of metaphysical generation between objects, events, properties, property instances, or items of other categories. Perhaps the electrons and protons metaphysically generate the atom. I have no opposition to such claims. But on them I remain officially agnostic.

In this paper, I will discuss a wide variety of stuff: objects, property types, property tokens, laws, events, and other material as well. It will be helpful to have a neutral term that can include items from any category. I choose the term 'phenomenon'. This seems appropriate. The world contains a tremendous variety of phenomena. Some are straightforwardly physical (leptons and mass); some are not (experiences of red, laws of a nation). *Fundamentality physicalism* says that all of these phenomena, of whatever type, ultimately depend on and are generated from the physical.

1.1.4 The Covering Constraint

The layout question asks, "What tokens of fundamentalia are there, and how are they?" It asks how the building blocks of reality are arranged. A complete answer to the layout question will tell us how every particular fundamental is. The complete layout of fundamentalia has an important feature, expressed in *the covering constraint*:

- **The Covering Constraint:** The complete layout of fundamentalia metaphysically generates all the world's contents.³

³Schaffer (2010): 38-9 discusses a similar constraint, which he titles "the tiling constraint".

The covering constraint entails that everything is either a fundamental or generated by some fundamental(s). The fundamentalia cover all of reality. I take the covering constraint to be constitutive of the notion of fundamentality.

I take claims about the fundamental to have modal import. This is an extra claim that is, strictly speaking, separable from the bare notion of fundamentality itself. But I endorse the modal import of fundamentality, and discussing it will both help the reader grip the notion of fundamentality and to make connections to extant philosophical theorizing about fundamentality, which has often been couched in the language of modality. The modality of which I speak is *metaphysical modality*, which I take to represent all the different ways a (possible) world could be. I take metaphysical possibility to be more permissive than physical possibility, mostly because I think that there are metaphysical possibilities (e.g. a different gravitational constant) that are not physically possible.

I said above that the layout of the world's fundamental constituents metaphysically generates the rest of the world's contents. Suppose that the world's fundamental level is composed of three types of particles - quarks, leptons, and bosons, some fundamental properties and relations, and laws that govern those particles, properties, and relations. On this assumption, once you fix the distribution of quarks, bosons, and leptons across space-time, assign them all their fundamental properties and relations, set in place the laws that govern their in-

Schaffer's tiling constraint involves two components. The first is equivalent to my covering constraint. The second is a "no overlap" clause. Schaffer claims that there are enough fundamentalia to cover all of reality, and no extras. Schaffer's "no overlap" clause makes more sense within the mereological framework within which he operates. However, I'm disinclined toward a no overlap constraint on the fundamental. I don't see why two fundamentalia couldn't do "unnecessary work", in the sense of one fundamental generating some non-fundamental that another fundamental also generates. A creator who builds a world with two fundamentals that overlap by doing the same metaphysical job creates a less elegant world than his rival creator who builds a similar world with less fundamentalia and no overlap. But sometimes the world is not elegant.

teraction, you thereby fix the entire contents of the world.

This talk of “fixing” can be made more specific using modal tools. The modal version of the covering constraint can be expressed as follows:

- **Modal Covering Constraint:** No two possible worlds differ without differing fundamentally.

Contrapositively, any two worlds that are fundamentally identical are completely identical. You can’t get an extra koala without adding some extra quarks, leptons, and bosons. Two worlds “differ fundamentally” when they differ in the layout, not necessarily the roster, of fundamentalia. Clearly, two worlds could differ with respect to the layout of fundamentalia (one world contains enough extra quarks, leptons, and bosons to generate an extra koala), without differing in their roster of fundamentalia. I will return to connections between fundamentality and modality in section 1.3 (“Modal Physicalism”).

1.1.5 Generation and Ground

I use the term ‘metaphysical generation’ (often just ‘generation’) to stand for the relation that holds between the fundamental and the non-fundamental. Fundamentalia generate non-fundamentalia. One might also use the term ‘ground’, declaring that the fundamental grounds the non-fundamental. The notion of ground has received much philosophical attention lately.⁴ The metaphysical generation between the fundamental and the non-fundamental is similar to what many call ‘ground’. Both are transitive, asymmetric, and irreflexive relations of metaphysical dependence in which one relatum holds in virtue of the other. Some

⁴For a start, cf. Fine (2012, 2001), Schaffer (2010, 2009), Rosen (2010), and the essays in Correia & Schneider (2012).

may have in mind what I've called 'generation' when they talk of ground. However, I am open to the thought that there may be important differences between the two. I believe that generation is one type, or species, of ground. I avoid the term 'ground' partly to prevent unwanted associations the reader may have with the term.

Philosophers introduce the notion of ground using examples of philosophical "in virtue of" claims. Here are some examples:^{5,6}

1. A thing has its dispositions in virtue of its categorical features.
2. An action is wrong in virtue of the action's failure to maximize utility.
3. The legal facts hold in virtue of (non-legal) social facts.
4. A set exists in virtue of its members.
5. A conjunction is true in virtue of the truth of its conjuncts.

For now, ignore the truth or falsity of these claims. Each of these claims has been endorsed by more than a few philosophers. These "in virtue of" claims come from a wide variety of philosophical areas. Yet (most) theorists of ground take all these claims to deal with the same thing: *ground*. A criticism often levied against fans of ground is that the disparate relations expressed in (1)-(5) and elsewhere are actually not all instances of the same relation. Research on some generic relation of "ground" is thus a waste of time. I think this dialectic expresses a false dichotomy. Philosophical research on ground is useful if the relations

⁵Examples (1)-(3) occur in Rosen (2010). Instances of (4) and (5) appear in Fine (2012). (4) also appears in Schaffer (2009).

⁶These claims are all universal in form. We can make similar specific statements, about what it is in virtue of which this particular wine glass is fragile, a given murder is wrong, a particular law holds, $\{\emptyset\}$ exists, or 'It's raining and it's sunny' is true.

in (1)-(5) and elsewhere are similar enough, and share enough formal features, that studying them together makes sense. The merit of philosophical research on ground is mostly independent of whether the relations in (1)-(5) are the same relation (ground), different species of the ground genera, or unrelated relations that share the property of being relations of metaphysical dependence. However, when lumping together different but formally and metaphysically similar relations from disparate areas of philosophical inquiry, we must keep in mind the dissimilarities between them. I believe that metaphysical generation is similar enough to these other relations it can fruitfully be studied under the banner of 'ground'. However, in doing so we must keep track of the peculiar features of metaphysical generation, features that may not have application in every arena in which the term 'ground' has been brought to bear. Since I am interested only in metaphysical generation, I choose not to risk such mis-application by using the generic term 'ground'.

Here's an example in which metaphysical generation may come apart from ground. Consider the following claim: "Knowledge entails justification". This is an important (and perhaps deep) truth of epistemology. In virtue of what is this claim true? What are its grounds? I'd like to tell two stories. One involves metaphysical generation; the other does not. Both involve "ground", in some sense of that term. Here's the story about metaphysical generation. "Knowledge entails justification" is a universal claim. The world contains many states of knowing that P. All of those are also states of having justification for P. My state of knowing that the food truck on La Brea Avenue has the best tortas is metaphysically generated by some incredibly complicated arrangement of neural states in combination with causal relations to the world. Those neural states and causal relations

are, somewhere down the line, generated by a more complicated arrangement of fundamental sub-atomic particles, properties, relations, and laws. I'm also in a state of being justified in believing that the food truck on La Brea Avenue has the best tortas. This state is also generated, via a long and complicated chain, by some arrangement of fundamentalia. Knowledge entails justification is generated by the fact that every time the fundamentalia get arranged in a manner that generates knowledge, the fundamentalia also get arranged in a manner that generates justification. This is the sense in which "knowledge entails justification" is generated from the fundamental.

But there's another story about what is in virtue of which knowledge entails justification. Knowledge entails justification in virtue of the nature of knowledge. That's how knowledge is. It always requires justification. In fact, one might even think that every knowledge-generating arrangement of fundamentalia is also a justification-generating arrangement *because* of this truth about the nature of knowledge. If the arrangement weren't justification-generating, it would not be knowledge-generating.

I think that both stories offer legitimate answers to the question, "In virtue of what does knowledge entail justification?" But they are different kinds of answers. One is a *generative* answer that looks for the fundamentalia in virtue of which the claim is true. The other embodies a different kind of answer. (I'm not sure how exactly to categorize the type). Both answers seem to represent some type of metaphysical in virtue of, and some variety of grounding relation. Importantly, in this paper I'm interested in the type of ground that goes with the generation of the non-fundamental from the fundamental: metaphysical generation.

1.2 Fundamentality Physicalism

1.2.1 Fundamentality Physicalism

The roster question asks, “What types of fundamentalia are there?” What kinds of building blocks generate our world? This is a tremendously natural and intuitive question. Philosophers and scientists have offered various answers. One is *fundamentality deism*: there is one fundamental: God. Another is *fundamentality idealism*: all the fundamentalia are mental. Physicalism is another popular thesis, and the current frontrunner for the true theory of the world’s fundamental level. *Fundamentality physicalism* says that all the world’s fundamentalia are, in some sense or other, physical. I’ll take *fundamentality dualism* to be the thesis that the world’s roster of fundamentalia contains at least two types of things: physical and mental. (For brevity, I’ll sometimes omit the qualifier ‘fundamentality’ when context allows.)

Fundamentality dualism can take many forms. Potential mental fundamentalia include consciousness particles, mental properties (“represents that P”, “experiences pain”), or a fundamental law of the universe that characterizes relations between the mental and the non-mental.⁷ For the most part, the discovery of reality’s fundamentalia falls to physics, especially theoretical and particle physics. Physics seems to suggest something like physicalism (though this claim is not uncontroversial). It has not postulated any consciousness particles yet.

Of course, there are issues about deciding whether or not some candidate fundamental counts as “physical” (cf. Dowell (2006a,b), Stoljar (2009a,b), Wilson (2006), Worley (2006)). It’s difficult to give necessary and sufficient conditions for

⁷Such laws are called “psycho-physical” by Chalmers (2003a, 1996a).

physicality in other terms. (See Stoljar (2009b) for a variety of attempts). Some, e.g. Stoljar (2009b), claim that this result robs physicalism of its interest as a philosophical thesis. I disagree.⁸ I think we have a good, if rough and ready, grasp on what kinds of fundamentalia jeopardize fundamentality physicalism. Fundamental quarks, mass, and photons do not count against fundamentality physicalism. Fundamental experiences of red, representations that P, and consciousness particles do. We know the physical, and the non-physical, when we see it, even if we can't give conditions for (fundamental) physicality in other terms. I don't take the situation to be much different with our grasp of most concepts, including 'koala', 'automobile', and 'table'.

It's notoriously hard to give necessary and sufficient conditions for being a C for almost any concept C. The concept "physical fundamental" is no different.⁹ This approach might lead to borderline cases of physicality, and of the truth of physicalism itself. Suppose it's borderline whether some property F counts as physical. If F is fundamental, then it will be borderline whether fundamentality physicalism is true. I don't take this to be a problem. Of course, what matters in the end is what the roster of fundamentalia is - what the fundamentalia are - and not whether those fundamentalia count as physical, mental or what-have-you. However, labels such as (fundamentality) 'physicalism', 'dualism', and 'idealism' are useful for characterizing broad classes of theories about the fundamental nature of our world, categorized by the roster of fundamentalia each theory posits.

Fundamentality physicalism says that all the actual world's fundamentalia are physical. I take fundamentality physicalism to constitute at least one reasonable

⁸For more on this disagreement between Stoljar and I, cf. Rabin (2011b).

⁹Cf. my Rabin (2011b) for more on this type of approach to defining 'physicalism' and 'physical'.

interpretation of the thesis that has gone by the names 'physicalism' and 'materialism' throughout philosophy's long history. We'll see some evidence for this claim in section 1.3 ("Modal Physicalism"), in which I'll argue that the modal definitions of physicalism that have been popular over the last thirty years derive their plausibility as a conception of physicalism from fundamentality physicalism.

1.2.2 Fundamentality Physicalism and the Mathematical

What is the relation between the physical fundamentalia and the realm of mathematics? The covering constraint says that everything is either fundamental or generated from the fundamental. Thus mathematical states of affairs, objects, and properties must be either fundamental or generated from the fundamental. If fundamentality physicalism is true, then mathematics must be either physical and fundamental or generated from physical fundamentalia. Both options are implausible.

Consider the mathematical state of affairs of 3 being prime. If mathematics is generated from physical fundamentalia like quarks and leptons, they must be generated in quite a different manner than my desk is generated. We can alter my desk by modifying the underlying fundamentalia that generate it. But no amount of re-arranging physical fundamentalia will alter the primeness of 3. This is evidence that the number 3 does not metaphysically depend on physical fundamentalia. In fact, given that the number 3 has its mathematical properties necessarily, it appears to be completely modally independent of the physical. Of course, these modal tests are defeasible evidence for metaphysical dependence and generation. But these tests tell against the dependence of the mathematical on the physical. The best candidate for the metaphysical generators of 3's prime-

ness are the axioms of arithmetic. But physical fundamentalia do no better in generating these axioms than they did in generating 3's primeness.

The nominalistically inclined may readily accept that the mathematical is generated from the physical. However, I wish to set aside that view for now.

Suppose that the mathematical is fundamental. Does this challenge fundamentality physicalism? This depends on whether fundamental mathematics should count as "physical". I wish to make two points against mathematical fundamentalia counting as "physical". First, fundamental mathematics would be quite different from the clearly physical quarks, leptons, and bosons, or masses, charges, and spins. It's also different from fundamental physical laws. Second, counting fundamental mathematics as physical has an odd result. Since mathematics is necessary, every possible world will contain physical fundamentalia.

Fundamentality physicalism is a contingent thesis about the fundamental structure of the actual world. While all physicalists should accept that fundamentality physicalism is true, few will accept that it is necessarily true. Physicalists tend to believe that (fundamentality) dualism is not actually true, but not that dualism is impossible. Fundamentality physicalism is compatible with the possibility, but not the actuality, of non-physical fundamentals. Fundamentality physicalism can accept the possibility of fundamental non-physical ether, spirits, ESP, mind substances, and all manner of physical anathema, as long as they are not actual.

Consider a possible world *Casper* whose fundamental level is purely spirit substance. Arrangements of spirit substance metaphysically generate the ghosts, ghouls, angels, and demons that inhabit *Casper*. *Casper* appears to contain no physical stuff whatsoever. But the truths of mathematics also hold at this world, and any theorist of this world's fundamental level will need to say something

about the mathematical. Suppose, as before, that the mathematical is fundamental. If we declare the mathematical fundamentals of the actual world to be physical, then it will turn out that *Casper*, an apparently purely spiritual world, contains physical elements at the fundamental level. This is an odd result. Furthermore, the same line of reasoning that led to declaring fundamental mathematics as “physical” can be run at Casper. The spirit philosophers of Casper have just as much reason to declare the mathematical as “spiritual”. If they’re correct, then the actual world contains fundamental spiritual elements! Perhaps these mathematical fundamentalia are both physical and spiritual? I’m not sure what to say.

If the mathematical is fundamental, it is fundamental at every world. I think that the best thing to say about the mathematical fundamentalia is that they are, in a certain sense, neutral in relation to debates over fundamentality physicalism, spiritualism, deism, or dualism. There are at least two ways to get this neutrality. First, we might admit that there are mathematical fundamentalia, and that these fundamentalia are neither physical, nor spiritual, nor etc. They are mathematical. Second, we might maintain that mathematics lies outside the scope of inquiry about the fundamental. Mathematics is a necessary tool for describing the world’s fundamental level and in some sense governs the behavior of the universe. But metaphysical generation does not apply to the mathematical. In either case, mathematics will not be physical, or spiritual, but it will be physicalistically acceptable, and spritualistically acceptable, in the sense of not challenging either fundamentality physicalism or fundamentality spiritualism.

Suppose we take the first option and declare mathematics fundamental and non-physical. In a certain sense, fundamentality physicalism is thereby falsified. But this result tells against the letter of fundamentality physicalism, not the spirit.

In the domains in which the nature of the fundamental is debated, such as the physicalism-dualism debate, it's clear that mathematics has been, and should be, set to the side. (Schiffer (1989) and Stoljar (1996) agree). Let's dub the thesis that all the fundamentalia are physical, no exceptions, *strong fundamentality physicalism*. This view requires either (i) that there are no mathematical facts (some type of nihilism or nominalism), (ii) that the mathematical facts are generated from the physical, or (iii) that there are mathematical fundamentalia which are physical. Let *weak fundamentality physicalism* be the thesis that all the fundamentalia are physical, bracketing all issues about mathematics.¹⁰

Weak fundamentality physicalism better captures the debate over the nature of the fundamental that has taken place under the title 'the physicalism-dualism debate'. Henceforth, when I use the term 'fundamentality physicalism' I will be speaking of weak fundamentality physicalism.

1.3 Modal Physicalism

In recent times, physicalism/materialism has often been categorized in modal terms. Modal definitions appear in Chalmers (1996a), Leuenberger (2008), Lewis (1983), Jackson (1993), Stoljar (2009b,a), Papineau (2002), and elsewhere. For our purposes, the differences between these formulations are minimal. We'll use Jackson's here.

Modal Physicalism: Every possible world that is a minimal physical duplicate of the actual world is a duplicate simpliciter.

The idea behind modal physicalism is simple: if you copy the actual world in all its physical respects, and don't add any extras, you copy the actual world in

¹⁰It's possible that additional issues must be set aside, including the place of abstracta and moral properties. I won't address these issues here.

all its respects. The physical is sufficient to yield the rest. A key idea here is that if two worlds were physically identical, but differed mentally (e.g. the first world contained a red experience where the second contained a blue experience), then it wouldn't be true that the physical was sufficient to generate the rest. More would be needed to fix the world's entire contents (e.g. the experiential contents). And physicalism, at least in some sense, would be falsified.

Two features of Modal Physicalism require clarification. The first is the notion of a *physical* duplicate. The second is the notion of a *minimal* duplicate. What does it take for a world to count as a physical duplicate of the actual world? Physicalists think that, in some sense or other, everything is physical, including mental states. But surely Jackson can't have meant to include mental states among the physicals that must appear at every physical duplicate of the actual world. That would prejudice the issue against the dualist and trivialize the definition. Modal physicalism says that copying all the physical stuff, in some restricted sense of 'physical', is sufficient to copy all the other stuff, including the mental (in the absence of extras). If we start by the counting all the "other stuff" as physical, the issue of physicalism vs. dualism is decided from the start.

It's clear enough what Jackson intends. If we copy all the actual world's electrons, leptons, bosons, molecules, masses, charges, spins, physical laws, and the like, we create a physical duplicate of the actual world. Let's call the material that must be duplicated in order to yield a physical duplicate of the actual world *strictly physical*. The physical duplicates of the actual world are the strictly physical duplicates.

We must also clarify the notion of a *minimal* duplicate. Before exploring the technicalia surrounding what it is to be a "minimal duplicate", it will be helpful to

describe the motivation behind appeal to the notion. Modal physicalism attempts to capture the test described above. Copy the world in all its physical respects. Add no extras. Modal physicalism is true if and only if in doing so you copy the world in all its respects.

Consider the following simpler definition of modal physicalism. *Simple modal physicalism* is true if and only if every (strictly) physical duplicate of the actual world is a duplicate simpliciter. Unfortunately, simple modal physicalism is incompatible with metaphysical possibilities with which many philosophers have thought physicalism should be compatible. Simple modal physicalism fails to capture to “copy the physical and add no extras” test.

Consider the world *Actual-plus-Ghost*. *Actual-plus-Ghost* is a physical duplicate of the actual world that has an additional element: a ghost that travels through space-time observing, but never interacting with, the physical. Physicalism about *Actual-plus-Ghost* is clearly false. That world contains non-physical ghosts. But, intuitively, the possibility of *Actual-plus-Ghost* should not challenge physicalism. (Or, at least, one coherent and substantive conception of physicalism, which modal physicalism aims to capture, is compatible with the possibility of *Actual-plus-Ghost*).

The metaphysical possibility of *Actual-plus-Ghost* falsifies simple modal physicalism. *Actual-plus-Ghost* is a physical duplicate of the actual world but not a complete duplicate. But *Actual-plus-Ghost* is not a *minimal* physical duplicate of the actual world, because it contains a non-physical extra: the ghost. Thus the possibility of *Actual-plus-Ghost* does not challenge modal physicalism defined using the “minimal duplicates” clause.

Physicalism is a claim about the metaphysical structure of the actual world. It claims that dualism is false, but not that dualism is metaphysically impossible. Physicalists believe Descartes was wrong about the existence of mental substances. But they need not think that Descartes imagined a metaphysical possibility. Some world might be the way Descartes thought the actual world to be. The possibility of worlds at which dualism is true, such as *Actual-plus-Astral*, does not challenge the physicalism that modal physicalists sought to capture. (Though some stronger forms of physicalism might deny such possibilities).

While the implications of modal physicalism about the actual world do not reach as far as *Actual-plus-Ghost*, they do reach other regions of modal space. For example, modal physicalism is incompatible with the metaphysical possibility of zombie worlds. A *zombie world* is a physical duplicate of the actual world (with no extras) that lacks all conscious experiences. *Zombie worlds* contains zombie doppelgangers of every actual person. These doppelgangers stroll the waterfront and utter sentences such as, "My, what a brilliant shade of blue." But they do not have conscious experiences of blue. The possibility of zombie worlds entails that the actual physicalia are modally insufficient to yield the actual conscious experiences. Physicalism would be, in at some important sense, false. Or at least this is one of the principle motivations behind modal formulations of physicalism.

Let's summarize. Philosophers searching for a modal definition of physicalism wanted a modal thesis compatible with the possibility (but not actuality) of *Actual-plus-Ghost*, but incompatible with the possibility of zombie worlds. They wanted a modal thesis that captured the "copy the physical and add no extras" test. Modal physicalism, defined using the "minimal physical duplicates" clause, accomplishes these goals. However, Jackson's use of minimal physical duplicates

is not the only method that will meet this desiderata. Chalmers (1996a) uses the notion of a *positive fact*, Leuenberger (2008) uses *ceteris absentibus modal sufficiency*, Lewis (1983) uses *alien properties*, and Stoljar (2009a) mentions a variety of possible strategies. The differences between these methods won't concern us here. We're primarily interested in the general strategy of using modality to define physicalism. Jackson's definition provides a useful paradigm.

We now move on to consider some of the technicalia associated with precisifying the notion of a minimal physical duplicate. Jackson's modal definition of physicalism presupposes a partial ordering \leq among the physical duplicates of the actual world. \leq must be reflexive, antisymmetric (i.e. $x \leq y$ and $y \leq x$ entail $x = y$), and transitive. Modal physicalism requires that all the minimal physical duplicates be duplicates simpliciter. Jackson does not say much about how the ordering should be established. Following Leuenberger (2008), there are at least two options.¹¹

\leq_F : $w \leq_F w_0 =_{df}$ every fundamental fact of w holds in w_0 .

\leq_P : $w \leq_P w_0 =_{df}$ every positive fact (fundamental or not) of w holds in w_0 .

Both orderings appeal to the notion of a *fact*. I plan to leave the notion intuitive. One might precisify using states of affairs or propositions. $w \leq_F w_0$ requires that the fundamental facts of w also hold at w_0 , but does not require the converse. w_0 might have more fundamentalia than w . \leq_P relies on the notion of a *positive fact*. Chalmers (1996a) defines a *positive fact of world w* as a fact that holds in every world that contains w as a proper part (40). This approach requires an understanding of what it is for one world to contain another as a proper part. Positive

¹¹My discussion of modal physicalism is indebted to Leuenberger (2008) and Chalmers (1996a): 38-42.

facts include the fact that my desk has the dimensions that it does, the fact that traffic in downtown Los Angeles is horrendous, and the fact that the Marianas Trench is more than 2 miles deep. Negative facts include explicitly negative facts as well as universally quantified facts. Examples include the fact that there are no ghosts and the fact that all humans were born on Earth.

All this talk about minimal duplicates and complete duplicates sounds more complicated than it is. Let's assume, quite plausibly, that there are no two distinct possible worlds that are complete clones. If modal physicalism is true, then there is one and only one minimal physical duplicate of the actual world: the actual world itself. If modal physicalism is false, there is either more than one such duplicate, or there is a physical duplicate of the actual world that is *more* minimal than the actual world. How might each of these occur? First, a zombie world - a physical duplicate of the actual world with no experiences whatsoever - would be a physical duplicate of the actual world that is more minimal than the actual world. Thus all the minimal physical duplicates of the actual world (i.e. the zombie world) would not be complete duplicates. Second, suppose that zombie worlds are impossible, but inverted qualia worlds are. An *inverted qualia world* is a physical duplicate of the actual world in which the experiences are switched around. When you experience blue, your inverted twin experiences red. When you experience a pain your twin experiences a tickle. In such a situation there would be multiple minimal physical duplicates of the actual world (all the various inverted worlds, including the actual world), but they would not all be complete duplicates of the actual world.

Modal physicalism is compatible with the metaphysical possibility of *Actual-plus-Ghost* because *Actual-plus-Ghost* is not a minimal physical duplicate of the ac-

tual world. On either interpretation of the ordering \leq on which modal physicalism relies, w -actual \leq *Actual-plus-Ghost*. *Actual-plus-Ghost* clearly contains more fundamentals than the actual world. Thus w -actual \leq_F *Actual-plus-Ghost*. *Actual-plus-Ghost* also contains positive facts that the actual world does not, such as the fact that ghosts exist. Thus w -actual \leq_P *Actual-plus-Ghost*.

Modal physicalism is a useful definition of physicalism. It gets a lot of things right. It is compatible with the possibility of *Actual-plus-Ghost*. It is incompatible with the possibility of zombie worlds and inverted qualia worlds. And, importantly, modal physicalism uses the familiar and well-understood tools of modality.

1.3.1 Two Criticisms of Modal Formulations of Physicalism

I have two major criticisms of modal physicalism as a definition of physicalism. First, the truth of modal physicalism guarantees only a modal covariation between the physical and everything else (including the mental). One might think that mere covariation is not strong enough for physicalism itself. Physicalism requires at least a dependence relation of the mental on the physical.¹² ¹³ Second, modal physicalism appears to yield the wrong verdict in certain cases.

I'll sketch a situation in which modal physicalism is true but intuitively physicalism is not. Russell (1927) imagined a situation in which the fundamental physical properties had underlying categorical bases. Deeper properties or entities make the electrons, bosons, and leptons have the charges and masses they do. Suppose that the underlying basis of the electron is a consciousness particle: a

¹²Modal covariation marks a type of modal "dependence". I claim that physicalism requires something stronger than the mere modal covariational form of "dependence".

¹³A similar criticism appears in Wilson (2005).

miniature experience. Assume also that, necessarily, the state of the consciousness particles necessitates the state of the electrons and vice versa. Each necessitates the other. Otherwise the fundamental level is purely physical. Under these assumptions, modal physicalism is true. Any world with the same physical state as the actual world will have the same electron state, which guarantees the same state of the consciousness particles. But in such a situation physicalism is clearly false. Physicalism, however one conceives it, is incompatible with consciousness particles in the bedrock of reality.

1.3.2 Fundamentality Physicalism and Modal Physicalism

Modal physicalism's miscategorization of the consciousness particle case points the way to a better definition of physicalism: fundamentality physicalism. What modal physicalism lacks is a dependence of the mental on the physical. In fact, I think that much of modal physicalism's plausibility derives from its affinity to fundamentality physicalism. Modal physicalism is as close as you can get to fundamentality physicalism without mentioning the words 'fundamentality', 'generation', 'dependence', or 'ground'. (In fact, if we precisify modal physicalism using \leq_F as the ordering that determines minimality, then modal physicalism already relies on a notion of fundamentality). I'd like to make three points in support of the claim that modal physicalism derives its plausibility from fundamentality physicalism. First, fundamentality physicalism entails modal physicalism. Second, modal physicalism is often motivated using the metaphor of God creating the world. This metaphor is just the creation test from section 1.1.1, which is a test for fundamentality. Third, when philosophers discuss cases in which modal physicalism is falsified, without fail discuss cases in which there are non-physical fundamentalia.

First, fundamentality physicalism entails modal physicalism. Modal physicalism requires that all the minimal physical duplicates be complete duplicates. Any world that is a physical duplicate of the actual world will have the same physical fundamentalia. Any world that is a minimal duplicate will contain no extra fundamentalia. Suppose fundamentality physicalism is true: all the actual fundamentalia are physical. Thus all the minimal physical duplicates of the actual world will be fundamental duplicates. By the covering constraint, all the fundamental duplicates are complete duplicates. Thus all the minimal physical duplicates are complete duplicates, i.e. modal physicalism is true. In fact, modal physicalism's clause about "minimal duplicates" is actually a very clever way of capturing the covering constraint using only modal tools.¹⁴

Second, modal definitions are often motivated as construals of physicalism using the metaphor of God constructing the world. The metaphor is borderline ubiquitous in the literature on defining physicalism modally. This type of reasoning appears in Leuenberger (2008): 1, Jackson (1993), and Chalmers (1996a): 38-41. These philosophers ask us to imagine God creating the world. If God need only create physical stuff to generate all of reality, then physicalism is true. If the physical (in the absence of extras) necessitates the rest, then "all God needs to create" is the physical. Of course, the God creation metaphor is exactly the creation test from section 1.1.1 ("The Building Blocks of Reality"). This test is most naturally construed as a test for what is fundamental. If God need only create the Xs to create the entire world, then that's good evidence that the fundamental building blocks of reality lie among the Xs.

¹⁴In fact, I believe that modal definitions of physicalism must appeal to the notion of fundamentality in order to solve the problem of being compatible with possibilities like *Actual-plus-Astral* but incompatible with the possibility of zombies and inverted qualia. I hope to argue this point in further work.

Third, when philosophers discuss situations in which modal physicalism is false, they discuss, almost without fail, scenarios in which fundamentality physicalism is false. Chalmers (1996a) considers two main hypotheses that would explain the failure of modal physicalism. First, he considers the possibility that there is a psycho-physical law of the universe that relate the mental and the physical (or the functional). This law is intended to be on a par with the fundamental laws of physics, such as the law of gravity. Second, he considers the possibility of a further non-physical layer of entities underlying the fundamental physical particles. These are clearly both scenarios in which fundamentality physicalism is false.

All of these considerations point toward the conclusion that modal physicalism, and philosopher's use of modal physicalism, is really just an attempt to get at the real McCoy: fundamentality physicalism. I don't take this point to be original. Others have made it recently, though they tend to use the lexicon of "ground"). For example, Schaffer (2009, 2003) writes that ground is *the* notion needed to define physicalism. Kim (1982) writes that modal covariation might be explained by a relation of metaphysical dependence such as metaphysical generation or ground:

Supervenience itself is not an explanatory relation. It is not a "deep" metaphysical relation; rather, it is a "surface" relation that reports a pattern of property covariation, suggesting the presence of an interesting dependency relation that might explain it. (Kim (1982): 167)

Modal physicalism is a claim about a certain type of supervenience on the physical. It's unsurprising that philosophers would reach for modal tools when attempting to characterize fundamentality physicalism. Modality is a familiar

notion with a well-understood formal framework. Metaphysical generation and ground are not.¹⁵ In fact, Lewis (1983): 29 lauds supervenience as “a stripped-down form of reductionism, unencumbered by dubious denials of existence, claims of ontological priority, or claims of translatability”. Modal formulations of physicalism do have some advantages. First, unlike fundamentality centered conceptions, modal formulations completely avoid the issues I addressed in section 1.2.2 about the relation between physical fundamentalia and mathematics. My strategy was to set these issues aside as irrelevant to the main locus of debate: the relation between the mental and the physical. But modal physicalism does not require any strategy at all. It neatly sidesteps the issue because these necessary existents and truths are trivially necessitated by anything. As Lewis points out, modal formulations are less committal because they don’t entail any dependence of the mental on the physical. Unfortunately, this lack of commitment means that modal definitions can fail to capture the intuitive notion of physicalism they sought to capture in the first place.

Lewis would praise modal formulations for their lack of “dubious” claim of ontological priority. I contest the “dubious” status of such claims. They are perfectly respectable. This essay does some work to explain what is meant by them. Fundamentality and metaphysical dependence is what we were after all along. As a definition of physicalism, fundamentality physicalism is preferable to modal physicalism. Fundamentality provides a better lens through which to view the physicalism-dualism debate. Even if a fundamentality-centered conception of physicalism is more risky than its modal brethren, the risks are worth the reward. (In Chapter 2: “Modal Rationalism and Two Replies to the Master Argu-

¹⁵The situation is changing as the formal and logical properties of ground receive more attention. Cf. Fine (2011), Rabin & Rabern (ms), deRosset (ms).

ment Against Physicalism”, I’ll argue that a fundamentality-oriented conception of physicalism allows us to see more clearly the key issues around which many arguments in the contemporary physicalism-dualism debate revolve.)

1.4 Other Physicalisms

1.4.1 Everything is Narrowly or Broadly Physical: A Schema

In its most general form, physicalism says that everything is physical. The thesis then requires clarification in two ways. First, what does the ‘everything’ range over? Second, what does it take for something to count as “physical”? Versions of physicalism will vary over their answers to these two clarificatory questions.¹⁶

If physicalism is true, then a wide variety of *prima facie* non-physical phenomena, including consciousness, representation, laws, and morality, must either be placed outside the scope of ‘everything’ or be counted as “physical” on some reasonable interpretation of that term. Often, a phenomenon will be counted as physical if it bears some chosen relation to more straightforwardly physical phenomena such as electrons, gravity, and the laws of thermodynamics. It is helpful to introduce the distinction between the *broadly physical* and the *narrowly physical*. The *narrowly physical* are physical in the more restricted sense of the term. Often, a conception of physicalism will count as narrowly physical all and only the phenomena studied by physics. Something is *broadly physical* if it bears the chosen relation R to the narrowly physical. Examples of such relations include metaphysical dependence and definability. A phenomenon can be broadly physical without being narrowly physical. If physicalism is true, the mental falls into this category.

¹⁶This framework is used by Stoljar (2009b).

Physicalism claims that everything is either broadly or narrowly physical. Different versions of physicalism will vary on (i) the range of ‘everything’ (ii) what counts as narrowly physical and (iii) the relation R something must bear to the narrowly physical to count as broadly physical. This framework will help us categorize a wide variety of formulations of physicalism, including fundamentality physicalism (section 1.4.2), modal physicalism (1.4.3), type identity physicalism (1.4.4), realization physicalism (1.4.5), nature physicalism (1.4.6), definition physicalism (1.4.8), and concept physicalism (1.4.8).

1.4.2 Applying the Schema to Fundamentality Physicalism

Strong fundamentality physicalism says that the quantifier ‘everything’ is unrestricted. It ranges over absolutely everything. Weak fundamentality physicalism places certain items outside its scope, including mathematics. Fundamentality physicalism counts something as *narrowly physical*_{fundamentality} if it is a physical fundamental. Something is *broadly physical*_{fundamentality} if it is metaphysically generated by narrowly physical fundamentalia. Fundamentality physicalism says that everything is physical, i.e. everything is either an narrowly physical fundamental or metaphysically generated from those fundamentals.¹⁷

Fundamentality physicalism has little bite without an understanding of what it takes for something to be “narrowly physical”. First, it must be fundamental. This much is clear. But what does it take for an alleged fundamental to count as “physical”? I won’t offer any independent criterion for the physicality of an alleged fundamental. We don’t need one. We know them when we see them. I find

¹⁷For completeness’s sake, I mention two other options. First, one might abandon my notion of “metaphysical generation” and use ground, or metaphysical/ontological dependence, instead. I’m not sure much turns on this. I use ‘metaphysical generation’ as label for whatever the metaphysical “making so” relation between the fundamental and the non-fundamental is. I take this relation to be a relation of metaphysical dependence.

the following heuristic useful: “If x were fundamental, would fundamentality physicalism be in trouble?” If the answer is ‘no’, then x is narrowly physical (if fundamental). If the answer is ‘yes’, x is not. Quarks, leptons, gravity, electrical fields, carbon, and weak nuclear forces all yield a ‘no’ answer. Not all of these are fundamental. But if they were, fundamentality physicalism would not thereby be challenged. Consciousness and representations yield a ‘yes’ answer. If either of these were fundamental then fundamentality physicalism would be in trouble.

Fundamentality is, in a certain sense, stingier than other forms of physicalism when it comes to the narrowly physical. Most other forms would count thermodynamical phenomena, such as an ideal gas law, as narrowly physical. Such phenomena provide paradigmatic subjects of physical inquiry. They appear in almost all introductory physics textbooks. But fundamentality physicalism counts thermodynamics as broadly, and not narrowly, physical, because thermodynamical phenomena are not fundamental. From fundamentality physicalism’s standpoint, thermodynamical and psychological phenomena are on a par. Both are generated from the narrowly physical fundamental level. However, fundamentality physical does recognize a difference between the two. If thermodynamical phenomena were fundamental, fundamentality physicalism would remain unscathed. If psychological phenomena were fundamental, fundamentality physicalism would be false.

1.4.3 Applying the Schema to Modal Physicalism

Modal physicalism can also be understood using the everything / broad / narrow schema. There is likely more than one way to fit modal physicalism into this schema. I’ll explain one such way. First we introduce the notion of a fact, which I will leave intuitive. A fact is simply some way the world is. ‘Everything’

ranges over all facts. Two worlds are complete duplicates if the same facts are true at both of them. The narrowly physical facts are the strictly physical facts I introduced in section 1.3 (“Modal Physicalism”). These are the facts about narrowly physical stuff: electrons, spins, weak nuclear forces, and the like. A fact is *broadly physical*_{modal} if it is *minimally necessitated* by all the actual narrowly physical facts. A fact is *minimally necessitated* by the actual narrowly physical facts if it holds at all the minimal physical duplicates of the actual world. Translated into this vocabulary, modal physicalism says that all the facts are broadly physical facts. (Modal physicalism, unlike fundamentality physicalism, counts all the narrowly physical facts as broadly physical, because every fact necessitates itself. Fundamentality physicalism uses metaphysical generation or dependence over necessitation, and nothing metaphysically generates or depends on itself.)

1.4.4 Type Identity Physicalism

Type identity physicalism is a version of the classic materialist identity theory in the philosophy of mind. Type identity physicalism is a claim about properties. ‘Everything’ ranges over properties. The narrowly physical properties are the properties studied in the physical sciences: charge, mass, being a lepton, having such-and-such spin, having such-and-such velocity in xyz direction, etc. A property is broadly physical if it is identical to some narrowly physical property. Thus the broadly physical properties are all and only the narrowly physical properties. Type identity physicalism says that every property is identical to some narrowly physical property.

According to type identity physicalism, despite appearances to the contrary, properties like experiencing red, having a good credit rating, and having a younger sibling are, in fact, narrowly physical properties studied by the physical sciences.

The physical sciences do not study those properties under those names, but they study them nonetheless. Perhaps the property of having a good credit rating is identical to the property that is the disjunction of all the various microphysical arrangements that yield the result that one has a good credit rating.

Type identity physicalism has often been objected to on the grounds that it is incompatible with the multiple realizability of the mental. The idea behind multiple realizability is that a mental state, such as a state of being in pain, might be realized by c-fiber firings in humans and by o-fiber firings in octopi. Identification of the property of being in pain with either the property of having c-fibers firing or the property of having o-fiber firings yields an intuitively incorrect result. Either humans or octopi don't experience pain.

One potential identity theorist reply has already been previewed. The type identity theorist could appeal to disjunctive properties. Identifying being in pain with the disjunction of having c-fiber and o-fiber firings avoids the problem. For this move to work, the type identity theorist will need to further identify properties such as having c-fibers firings with further (probably infinite) disjunctions of narrowly physical properties.

There are many moves to be made here, I won't go through the details. Perhaps the shift to disjunctions robs the type identity theory of some of its initial appeal. Or perhaps the disjunctive identity theory is too weak a form of physicalism for some. Or perhaps that fact that the physical sciences study property F and property G does not entail that it studies the property F-or-G.

1.4.5 Realization Physicalism

Melnyk (2003) proposes *Realization Physicalism*. Realization physicalism is a claim about token objects, properties, and events. According to realization physicalism,

a token object, property, or event is *narrowly physical*_{realization} if it is of a type studied by physics. A token object, property, event is *broadly physical*_{realization} if it is realized by some narrowly physical object, property, or event. Realization physicalism says that every token object, property, and event is either narrowly or broadly physical.

Roughly, *x realizes y* just in case *y* is a token of a functional type characterized by some condition *C* (such as playing a certain causal role) and *x* is a token of a type that satisfies that condition. A token electronic state of a laptop computer realizes a computational state because the latter is definable by reference to a functional role and the electronic state plays that role. Melnyk imagines a chain of realization from macroscopic objects, properties, and events we encounter every day, through medium-sized phenomena and eventually down to the ultimate realizers of fundamental physics.

Because of its focus on particular objects, properties, and events, realization physicalism is compatible with multiple realizability. This token of human pain can be realized by *c*-fiber firings (which are in turn realized by narrowly physical property P_1), while this token of octopus pain can be realized by *o*-fiber firings (in turn realized by narrowly physical property P_2).

Realization physicalism is similar to fundamentality physicalism. Both envision a layered picture of reality with an ultimate bedrock of physical fundamentalia. Melnyk (2003) even uses the term 'fundamental', and claims that the ultimate realizers of everything are physical fundamentalia. Realization physicalism, like fundamentality physicalism, entails modal physicalism (because realizers necessitate that which they realize). And realization physicalism, like fundamentality physicalism, and unlike modal physicalism, gets the intuitively cor-

rect verdict (physicalism is false) in the case from section ?? of the consciousness particle underlying and necessitating electrons.

There are two main differences between fundamentality physicalism and realization physicalism. First, they offer differing pictures of the relation between the physical bedrock and the higher-level world. My relation of metaphysical generation entails metaphysical or ontological dependence. Realization needn't have this commitment (though one might choose to add it or to argue that realization entails some type of dependence). Second, if the higher-level features of our world are realized by narrowly physical features, then these higher-level features must be in some sense functional. x realizes y entails that y is a token of a functional type. If realization physicalism is true, then all the non-narrowly physical objects, events, and properties are functional. This claim is, at the least, controversial. Fundamentality physicalism has no such commitment.

1.4.6 Nature Physicalism

Arguably, every thing in the world has a nature, or essence. The nature of a thing is revealed in the deep explanation of what that thing is. The nature of a thing can be captured by something like an Aristotelian *real definition*. A *real definition* of x explains what it is to be x . For example, the nature of water is to consist of molecules combining two hydrogen atoms and one oxygen. That's what it is to be water.

Nature physicalism, like other physicalisms, says that everything is physical. Mass, spin, and quantum entanglement are narrowly physical. x is *partly broadly physical*_{nature} if the nature of x involves narrowly physical phenomena. x is *wholly broadly physical*_{nature} if the nature of x involves only narrowly physical phenomena. Wholly broad physicality_{nature} entails partly broad physicality_{nature}. Whether

the nature of x involves y can be determined by whether y must be mentioned in the real definition, or deep explanation, of what it is to be x . For example, suppose conscious experience occurs because of a certain type of collapse of the quantum wave function. This type of collapse is key to understanding consciousness. If so, then the nature of conscious experience would be at least partly broadly physical.

Nature physicalists can say a variety of thing about the quantifier ‘everything’. They might choose to exclude mathematics, or morality, or abstracta. Or they might not. Less exclusion yields a stronger physicalism.

We get (at least) two versions of nature physicalism. *Whole nature physicalism* says that everything has a wholly physical nature. *Part nature physicalism* says that everything has a partly physical nature.

Examples of things with at least partly (and perhaps wholly) physical natures include electrons, black holes, the elements of the periodic table, and water. The truth of nature physicalism will turn on whether various macroscopic phenomena have (wholly or partly) physical natures. Items such as representations, consciousness, morality, and nations all pose challenges.

1.4.7 Explanatory Physicalism

Explanatory physicalism says that all the varieties of phenomena we encounter in the world can be understood using the tools and methods required to understand narrowly physical phenomena. ‘Everything’ ranges over phenomena that require explanation. *Narrowly physical_{explanatory}* phenomena are the phenomena studied by some specified range of physical sciences. Clearly physics is one such science. But the explanatory physicalist might choose to include chemistry, geology, or astronomy. For now, let’s interpret “narrowly physical_{explanatory}” generously, to include as much as seems reasonable. Certain phenomena should not be counted

narrowly physical, including the fluctuations of the stock market, the changing colors of cuttlefish, and the meaning of the word 'scrumptious'. But explanatory physicalism counts these phenomena broadly physical if they can be understood using the tools, methods, and concepts used in physics, chemistry, and the other physical sciences (whatever they may be). If they can all be so understood, explanatory physicalism is true.

Explanatory physicalism is similar to nature physicalism. Explanatory physicalism is an epistemic claim about explanation and understanding, while nature physicalism is a metaphysical claim about natures. However, the nature of x is revealed by the deep explanation of what it is to be x . Thus for both nature physicalism and explanatory physicalism, the (broad) physicality of a thing turns on what is needed to understand that thing. The entailment relations between the two will turn on the relation between (a) understanding/explaining in the sense required by explanatory physicalism and (b) capturing the nature in the sense required by nature physicalism. I can imagine views with every permutation of entailment between (a) and (b).

Explanatory physicalism, in its strongest form, maintains that all genuine explanations must be couched in the terms or concepts of physics. Explanation in other terms, while sometimes useful or pragmatic, will or should eventually be discarded as the explanatory scope of physics grows and engulfs the phenomena the non-physical sciences sought to explain. The deepest insight comes from explanations expressed in narrowly physical vocabulary.

While explanatory physicalism is similar in some ways to nature physicalism, explanatory physicalism has an advantage over nature physicalism. Explanatory physicalism can be endorsed as a definition of physicalism by those who are

skeptical that things have natures at all.

1.4.8 Definition Physicalism

Definition physicalism is a linguistic, rather than metaphysical, thesis. According to definition physicalism, 'everything' ranges over some specified class of expressions: singular terms, predicates, terms of the special sciences, complete sentences, etc. A term is *narrowly physical*_{definition} if and only if it appears in the physical sciences.¹⁸ A term is *broadly physical*_{definition} if and only if it can be translated into narrowly physical terms. Different versions of definition physicalism offer different criteria for when a translation or definition is successful. At the least, one should require that the extension of the definiens be the same as the definiendum. This entails that the definiens provide necessary and sufficient conditions for the definiendum. One might also maintain that that definiens have the same meaning as the definiendum. In what follows, I will not insist on this requirement.

Definition physicalism is a very bold thesis. I think it highly implausible, even when 'everything' is limited to a very narrow class of terms, such as biological, or even geographical, terms. It's tremendously difficult to correctly define any term using others. Dictionary definitions rarely accomplish this feat. Instead they facilitate understanding of the term, by pointing one in correct direction. Rarely do dictionaries give necessary and sufficient conditions. Furthermore, even when a definition succeeds (in the sense of providing necessary and sufficient conditions), it usually does so by utilizing cognate expressions that in turn can only be defined using the originally defined term. Given how hard it is to success-

¹⁸This definition will count many mathematical terms as narrowly physical_{definition}. I'm not sure whether proponents of a definition oriented conception of physicalism will welcome or resist this result. One could re-work narrowly physical_{definition} to count only non-mathematical terms.

fully define any term, I think that it highly implausible that any large or interesting class of terms can be defined using only the limited vocabulary of physics.

1.4.9 Concept Physicalism

Like definition physicalism, concept physicalism is more epistemic than metaphysical. Concept physicalism is a claim about the conditions under which an agent can possess, or master / fully understand, a concept. Before stating what concept physicalism is, we must introduce the distinction between concept possession and mastery. Concepts are the basic units of thought. The thought OSTRICHES LIKE CHOCOLATE is composed of the concepts OSTRICHES, LIKE, and CHOCOLATE. (Note on terminology: I use small capitals to express items at the conceptual level, e.g. the concept CHOCOLATE or the thought OSTRICHES LIKE CHOCOLATE.) An agent *possesses* a concept when he or she can think or consider thoughts of which the concept is a component. A thinker has *conceptual mastery* of a concept when he or she fully understands that concept.

Thinkers can possess concepts without having mastery. Burge (1979)'s ARTHRITIS case provides a classic example. Burge imagines Alf, who approaches his doctor complaining of arthritis in his thigh. Alf's doctor corrects him, pointing out that one cannot have arthritis in the thigh. Arthritis, by definition, affects only joints. Alf accepts the correction, admits that he was mistaken to believe that he had arthritis in his thigh, and modifies his usage of 'arthritis' and ARTHRITIS accordingly. The important point here is that Alf was able to think the thought I HAVE ARTHRITIS IN MY THIGH, and thereby possessed the concept ARTHRITIS, without fully understanding, i.e. mastering, that concept. Examples of this nature can be cooked up for almost any concept whatsoever. Many people use technical concepts that have worked their way into public consciousness with-

out mastering them. Examples include DARK MATTER, CHAOS THEORY, CLOUD COMPUTING, TRANSISTOR, ID, FRACKING, and HEDGE FUND.

Concept physicalism comes in four versions. It says that an agent who possesses/masters all the physical concepts is in a position to, through thinking alone, possess/master all concepts. Technically, concept physicalism comes in four varieties (possession-possession, possession-mastery, mastery-possession, master-mastery). We consider only two, primarily because these are the two most interesting theses.

- **Concept Possession Physicalism:** An agent who masters all physical concepts will be in a position to, through thinking alone, possess all concepts.
- **Concept Mastery Physicalism:** An agent who masters all physical concepts will be in a position to, through thinking alone, master all concepts.

The idea behind concept physicalism, of either variety, is simple. Take someone who fully understands all physical concepts. Allow her to think for as long as she wishes. Idealize past features including the thinker's fatigue, boredom, lifespan, and computational limits. For each concept, if she can come to possess that concept, then concept possession physicalism is true. If she can mastery each concept, then concept mastery physicalism is true.

Physical concepts are the concepts used in physics, or expressed by the terms of physics. These includes concepts such as ELECTRON, MASS, FORCE, and GRAVITY. One might also let mathematical and functional concepts count as physical.

The following thought, which I take to be fairly common, supports both forms of concept physicalism: *All concepts are possessable and fully understandable from the armchair.* If this is correct, then the truth of concept physicalism has nothing in

particular to do with the physical. If we start with a thinker who has no concepts at all, or perhaps some minimal stock of concepts required to get thought off the ground in the first place, she'll be able to possess and master all concepts. Another line of thought that supports concept physicalism is that all concepts are somehow formed, or constructed from, physical concepts. This thought lies along the same track as definition physicalism. It is in some ways stronger than definition physicalism, which makes no claims about "construction of concepts".

I think that both forms of concept physicalism are false. A variation on Jackson (1982)'s Mary thought experiment defeats them. Imagine Mary, a brilliant, but blind, scientist raised in a black and white room. Mary learns all the physical concepts and fully understands them. She publishes important results in solid-state physics. But Mary's colleagues, for whatever reason, never reveal to Mary that she is blind or in any way different than other people. Mary learns all about light-waves and their properties. But she never learns that humans have a capacity to perceive those light-waves using the sense of sight. Mary never experiences the sensation of seeing red. Consider the concept RED_{ph} . RED_{ph} is the phenomenal, experiential, concept of red. It applies to experiences or a certain type (you know the ones), and not to light-waves or to the surfaces of objects. Mary does not possess the concept RED_{ph} . If we place her in an armchair and allow to think until the end of time, she will not come to either possess or master the concept RED_{ph} . Therefore, both conception possession and concept mastery physicalism are false.

I have argued elsewhere (Rabin (2011a), Chapter 3) that Jackson's "knowledge argument" does not defeat modal physicalism. I am similarly doubtful that it defeats fundamentality physicalism (the shift from modality to fundamentality

does not significantly alter the dialectic). However, Jackson's thought experiment does defeat concept physicalism.

1.4.10 Summary

Each form of physicalism discussed says that everything is either broadly or narrowly physical. With the exception of type identity physicalism, each recognizes a difference between the narrowly and the broadly physical. Each form offers a different account of the relation R something must bear to the narrowly physical to count as broadly physical.

- **Fundamentality Physicalism:** x is broadly physical iff x is metaphysically generated from and dependent on the narrowly physical (which must be fundamental).
- **Modal Physicalism:** x is a broadly physical fact iff x is minimally necessitated by the narrowly physical facts.
- **Type Identity Physicalism:** x is a broadly physical type (property) if and only if x is a narrowly physical type (property).
- **Realization Physicalism:** x is a broadly physical token object, property, or event iff x is realized by some narrowly physical object, property, or event.
- **Nature Physicalism:** x is broadly physical iff the nature of x is partly/wholly narrowly physical.
- **Explanatory Physicalism:** x is a broadly physical phenomenon iff x can be understood using the tools and methods needed to understand narrowly physical phenomena.

- Definition Physicalism: A term x is broadly physical iff x is definable in narrowly physical terms.
- Concept Physicalism: A concept x is broadly physical iff x is possessable/masterable, through thinking alone, by someone who has mastered all narrowly physical concepts.

Lastly, it's worth mentioning that what I've offered is surely not a complete catalogue of the claims that have gone or will go under the names 'physicalism' or 'materialism'. Various other criteria might be offered for whether something counts as "physical", in the broad sense.

1.5 Fundamentality Physicalism and Other Physicalisms

1.5.1 Against Nature and Explanatory Physicalism

I think that whole nature physicalism and explanatory physicalism are both false. Many non-narrowly physical phenomena resist explanation, and understanding, without the use of concepts and tools specific to the subject of inquiry. Often, the science of that subject begins to flourish only once the appropriate concepts are introduced. These concepts allow for proper framing of the subject under investigation. For example, economic fluctuations can't be understood without the concepts of *supply* and *demand*. Representations can't be understood without some concept of *veridicality* or *accuracy*. *Truth* is one such concept.

The most basic function of a representation is to represent the world as being certain way. The sentence 'The taco truck on La Brea Avenue has the best tortas' says that the world is a place where the best tortas can be found at the taco truck on La Brea Avenue. When a representation succeeds at this function it is accurate. When it does not, it is inaccurate. Any account of representation that ignores

these features will be woefully incomplete. Furthermore, the notions of accuracy and inaccuracy cannot, in turn, be explained using the tools and methods of the physical sciences.

Thus, the nature of representation is, at least partly, not narrowly physical. That nature involves accuracy. This entails that whole nature physicalism is false. Similarly, the explanatory tools needed to understand representation include tools and concepts not found in the study of the narrowly physical. These include the concepts of accuracy and the tools of semantics. Thus explanatory physicalism is false as well.

I recognize that these claims are somewhat controversial. For example, teleosemantic approaches to meaning (Millikan (1989, 1993, 1984)) claim that representation can be understood using the tools and concepts of biology, specifically the notion of *biological function*. If biology can in turn be understood using the tools and concepts of physics, then representation can be similarly understood. I find all the steps in this alleged reduction of representation to physics dubious. But this is not the place to push the point. My goal is to sketch why I think these forms of physicalism too strong, and demonstrate that the points that tell against these stronger physicalism do not harm fundamentality physicalism.

1.5.2 The Relevance of Fundamentally Non-Physical Worlds

Recall the world Casper from section 1.2.2. Casper is a world whose fundamental layer consists of spirit stuff, including spirit particles called “ghostons” and “ectoplasmons” which float in an ether are governed by a force called “kismet”. At the non-fundamental level, Casper is by inhabited ghosts, ghouls, demons, and angels, all of whom have mental states. The fundamental bedrock of Casper consists entirely of spirit stuff. No quarks, no mass, no charge, no nuclear forces,

no gravity. I believe that worlds like Casper are metaphysically possible. I'd like to first argue for this claim and then draw out some of the lessons such worlds offer regarding physicalism.

When I speak of "worlds like Casper", I mean worlds that contain non-physical fundamentalia, and in which those non-physical fundamentalia metaphysically generate a rich variety of non-fundamentalia, including thinking agents.

1.5.3 Fundamentally Non-Physical Worlds are Metaphysically Possible

My main argument for the possibility of possible worlds with non-physical fundamentalia is that the denial of their possibility involves a type of parochialism about modality. Modality is not the place for parochialism. We should be liberal about what is metaphysically possible. If we can conceive it, we should take it to be possible unless we have evidence to the contrary. We can conceive worlds like Casper. There is no evidence for the impossibility. Therefore they are possible.

I will consider three possibilities for the denial of the metaphysical possibility of *Casper*. First, one could deny that there can be fundamental particles, properties, or forces, other than those instantiated at the actual world. Second, one could accept non-physical fundamentalia, but deny that these fundamental can generate macroscopic such as ghosts, ghouls, angels, and demons with mental states. Third, one could claim that the fundamentalia of Casper are physical.

Consider the objector who denies that there can be fundamental particles, properties, or forces, other than those instantiated at the actual world. This view entails that every possible variety of fundamental is actually instantiated. Here's an argument against the view. There are sub-regions of the actual world without electrons. Some of these sub-regions are identical to entire possible worlds. Therefore there are possible worlds without electrons, and not all worlds contain

every variety of fundamental stuff. Given that some worlds do not contain every variety of fundamental stuff, it is unmotivated and parochial to think that we happened (by dumb luck!) to inhabit one of the few worlds that contains every variety. Therefore, the actual world does not contain every possible type of fundamental.

Consider now the objector who admits the possibility of non-actual fundamentalia, but denies that these fundamentalia can metaphysically generate macroscopic phenomena like people, or angels, or experiences. This form of modal parochialism is just as bad as the previous. It maintains that there could be alien non-actual fundamental stuff, but that the actual world just happens to contain all the interesting fundamental stuff - the stuff capable of forming macroscopic phenomena like people and experiences.

Lastly, consider the objector who admits the possibility of Casper, but claims that Casper is entirely physical. I'm not sure exactly what to say against this view. It seems crazy to me. Descartes believed in non-physical mental substances. Let Casper have some of that in its fundamental bedrock. Does this objector also maintain that if Descartes were correct about the existence of mental substances, fundamentality physicalism would remain true? I begin to wonder what this objector means by 'physical'.

The preceding arguments are not likely to convince the dedicated opponent. She might readily accept the modal parochialism to which I've argued denial of the possibility of Casper commits one. However, from here on I'll take the metaphysical possibility of worlds like Casper, in which non-physical fundamentalia generate a rich variety of non-fundamentalia, as given.

1.5.4 Space and Time

It's not unreasonable to think that even Casper, a world whose fundamental building blocks are mental substances and ghoston particles, contains at least one physical fundamental: space-time. After all, I said earlier (section 1.2.2) that the inhabitants of Casper *move* from one area of the universe to another. If they do so Casper must have some type of space. If they make these movements over time, Casper has something like time.

There is something to the idea that any possible world, or perhaps any possible world with thinking agents, must have something like space and/or time. In this sense such worlds may never be completely non-physical. However, it may be possible for space and/or time to be non-fundamental, and metaphysically generated from non-physical fundamentalia.¹⁹ But if even if space and time are fundamental, even fundamental at every possible world, they seem to play a different role than, say, quarks, leptons, and mass. Space and time provide something like the canvas on which the fundamentalia can be placed. Quarks, leptons, and weak nuclear forces are the paint on the canvas. Perhaps every possible world requires a canvas on which the fundamentalia are placed. If so, perhaps every possible world contains something space-time-like. But this does not challenge the idea that non-physical fundamentalia can be placed on this canvas, or that the canvas can be mostly devoid of physical fundamentalia.

When I imagine strange possible worlds such as Casper, I imagine them containing something like space-time. I find it easy to imagine worlds without quarks

¹⁹Theoretical physicists seem to entertain the possibility that space-time itself is not fundamental, and is instead generated from something deeper. If so, space-time, like many non-fundamentalia, may be generatable from a wide variety of fundamentalia, including non-physical fundamentalia. Cf. Seiberg (1996).

or mass, but difficult to imagine worlds without space-time. Lewis (1986) thought similarly, claiming that each possible world was a spatio-temporally connected unit. I'm not sure whether this is the result of a genuine metaphysical restriction on the nature of possibilities, or a limitation due to the fact that my understanding and experience of the world is so strongly shaped by space-time itself.

Further muddying the waters is the difficulty of determining when some arbitrary 4-dimensional manifold should count as "a space-time". Need it have a directedness, in the way actual time does? It seems easy enough to assign non-physical fundamentalia to "locations" in the four dimensional manifold, and then establish laws for how those fundamentalia evolve in a phase space, of which the four-dimensional manifold provides four coordinates. This type of move can yield many of the same behaviors and complexity we'd want from a space-time "canvas" on which fundamentalia lie. But it's not clear to me what needs to be added, if anything, to make this four-dimensional manifold a genuine space-time.

Moving on, I'm willing to provisionally accept the idea that any possible world with interesting non-fundamentalia of the kinds we're accustomed to, including thinking agents, might also need to include at least one physical phenomenon: space-time.

1.5.5 Lessons from the Possibility of Fundamentally Non-Physical Worlds

The non-physical fundamentalia of Casper metaphysically generate ghosts, ghouls, angels, and demons with mental states. Casper contains many (non-fundamental) phenomena with which we are intimately familiar. The creatures of Casper form friendships, play games, establish nations, go to school, manipulate the stock market, and create works of art. They do all of this without quarks, leptons,

bosons, or any physical forces whatsoever.

The possibility of non-fundamental phenomena such as friendships and games that are metaphysically generated from non-physical fundamentalia should make us wary of strongly reductive theses such as type identity, nature, and definition physicalism. If 'friendship' can be defined in the narrowly physical terms of physics, then friendship cannot exist where the referents of those physical terms do not exist. But friendship does exist in such situations, including in Casper. Thus definition physicalism is false. A similar point applies to nature physicalism. If the nature of friendship involves quarks, or gravity, then friendship cannot exist without quarks, or gravity. But friendship can exist without any physical fundamentalia whatsoever (with the possible exception of space-time). This should make us suspicious that friendship has a physical nature at all. If friendship is identical to some narrowly physical type, friendship cannot occur at Casper. But it does. Thus friendship is not identical to any physical type. Type identity physicalism is false.

Similar points to a huge variety of non-fundamental phenomena. The examples I've given all require mentality, or representational capacities, in some way. But the point can be made using phenomena such as life, or computers, that don't depend on mentality and associated representational capacities. Casper could contain the ectoplasmic equivalent of amoeba: living creatures with no representational capacities. Or Casper could contain naturally occurring ectoplasmic computers.

I'd like to offer an alternative picture. A wide variety of the world's macroscopic phenomena are *neutral* regarding their underlying metaphysical generators. In order for friendship, games, representation, computers, spherically-

shaped objects, or life to occur, there must be some underlying substrate that metaphysically generates the phenomenon in question. But games, computer, and life care very little about the nature of the fundamentals that underly them. Furthermore, (this is an additional claim), many different types of fundamentalia, some physical, some non-physical, can do the job. To be a computer just is to be something with a variety of states in the appropriate very complicated relations. These relations can be instantiated by appropriate arrangements of quarks, leptons, and bosons. But they can also be instantiated by arrangements of ghostons and ectoplasm. The nature of a computer makes no reference to quarks or leptons. One can fully understand computers without knowing anything about the fundamental physical structure of the world. The explanation of computation uses a special vocabulary appropriate to that subject matter, a vocabulary that includes terms like 'program' and 'transistor'.

Not all the world's non-fundamental phenomena are neutral in this way. Wooden tables require wood, which requires carbon, which requires a certain combination of electrons and protons. Wooden tables are not neutral in the same way that mentality, life, and computers are. Water and steel provide similar examples.

These points are reminiscent of the idea of "multiple realizability" in the philosophy of mind. Minds, and other mental phenomena such as pain, are allegedly multiply realizable by a wide variety of underlying substrates, including brains and computers. I'm claiming that many non-fundamental phenomena, including minds and pains, are "multiply generatable". They can be generated by a wide variety of arrangements of fundamentalia, including completely non-physical fundamentalia. Most conceptions of multiple realizability require that the multiply realizable phenomenon have some type of topic-neutral functional

analysis or account. The phenomenon is then realized by whatever underlying stuff satisfies the functional analysis or account. Multiple generatability requires no such commitment to any type of functionalism.

Even if one denies the metaphysical possibility of worlds such as Casper, one can still accept the idea of multiple generatability.

1.5.6 Fundamentality Physicalism and the Fundamentally Non-Physical

Fundamentality physicalism is compatible with worlds such as Casper. Realization physicalism is as well. Each claims only that the all the actual fundamentalia or ultimate realizers are narrowly physical. Neither denies the possibility of non-physical fundamentalia or ultimate realizers.

Fundamentality physicalism, unlike stronger physicalisms, can accept the multiple generatability of many non-fundamental phenomena. Fundamentality physicalism says only that, of the actual instances of these phenomena, they are all generated from and metaphysically depend on narrowly physical fundamentalia. Unlike nature physicalism, fundamentality physicalism does not require that the nature of the mental be in some way related to or dependent the physical. Unlike explanatory physicalism, fundamentality physicalism can accept that the mental, the economic, and the living can be understood using a set of tools and concepts that are specific and idiosyncratic to the subject of inquiry, and which can't somehow be reduced to concepts or methods of the physical sciences. Fundamentality physicalism, unlike some other physicalisms, offers psychology, economics, and biology, a variety of explanatory autonomy from the physical sciences and physics itself.

Fundamentality physicalism does have mild reductive ambitions. It claims that everything metaphysically depends on the narrowly physical fundamen-

talia. All the actual psychological, economic, and biological phenomena and states of affairs are they way they are virtue of the way the physical fundamentalia are distributed. But this type of reductive approach stops far short of the stronger reductive aims of nature physicalism, explanatory physicalism, definition physicalism, or even functionalism in the philosophy of mind.

1.6 Conclusion

My main goal in this essay was to introduce fundamentality physicalism and its key concepts of *fundamentality* and *metaphysical generation*. A secondary goal was to introduce several other theses that could reasonably go under the names ‘physicalism’ and ‘materialism’, and to systematize them using the ideas of the *broadly* and the *narrowly* physical. Fundamentality physicalism is weaker than most of these other theses, including type identity, nature, explanatory, and definition physicalism. Fundamentality physicalism allows many of the non-narrowly physical features of our world a type of explanatory and metaphysical autonomy from the narrowly physical that these others forms of physicalism do not. This is a positive feature of fundamentality physicalism. It makes the thesis more likely to be true.

Foremost among these other conceptions of physicalism was modal physicalism. Modal physicalism certainly has been called ‘physicalism’ many times. It is the form of physicalism that has dominated the massive literature on physicalism, the physicalism-dualism debate, and the metaphysics of consciousness over the past 20 years.²⁰ However, I argued (section 1.3.2 “Fundamentality Physical-

²⁰The literature in this area is massive and I cannot possibly list it all or even the highlights. Useful resources include Braddon-Mitchell (2003), Chalmers (1996a, 2010, 1999, 2003a), Leuenberger (2008), Stoljar (2009b,a, 2001), Block & Stalnaker (1999), Chalmers & Jackson (2001), McGinn (1989), Papineau (2002), Loar (1990), Melnyk (2003), Hawthorne (2002), Levine (2004), Jackson

ism and Modal Physicalism”) that fundamentality physicalism more accurately describes what has actually been debated in that literature: whether some non-physical features, in particular mental features, are fundamental constituents of reality.

Fundamentality physicalism is a substantive thesis that is worthy of wearing the title ‘physicalism’. There are other physicalisms, but this form of physicalism is the relevant one for much of the contemporary debate.

(1982, 1986), Lewis (1988), Nida-Rumelin (2009, 1996),

Chapter 2

Modal Rationalism and the Demonstrative Reply to the Master Argument Against Physicalism

Chapter Abstract

According to the *Master Argument Against Physicalism*, an a priori gap between the physical and conscious experience entails a lack of necessitation and with it the falsity of physicalism. This paper investigates the crucial premise of the master argument: the inference from an a priori gap to a lack of necessitation. This premise gets its support from *modal rationalism*, according to which there are important, potentially constitutive, connections between a priori justification and metaphysical modality. I argue against the strong form of modal rationalism that underwrites the master argument, and suggest a more moderate rationalist view. I offer a novel *demonstrative reply* to the master argument, according to which demonstratives play a vital role in the generation of meaning for our linguistic and conceptual representations of conscious experience (such as 'pain'). I suggest that this connection between conscious experience and demonstratives, rather

than a metaphysical gap generated by the truth of dualism, is the source of the epistemic gap between consciousness and the physical.

2.1 Introduction

I believe that our world is, at its most fundamental level, physical. Call this thesis *fundamentality physicalism* (to which I will often refer using the simpler term ‘physicalism’). The entire world is the way it is in virtue of the way the world is at the microphysical level. All the cameras, cantaloupes, coffees, and koalas are the way they are in virtue of the way the leptons, bosons, and weak nuclear forces are. The basic building blocks of our world are purely physical. The rest of our world, including cameras, cantaloupes, and koalas, are generated from, and metaphysically depend on, the arrangement of the fundamental microphysical stuff.¹

In this paper I wish to address a certain contemporary challenge to physicalism as just described. I call this challenge *the master argument against physicalism*. The master argument starts from an epistemic a priori gap between the physical truths and the truths about phenomenal conscious experience.² The argument

¹Two points deserve clarification. First, physicalism defined in terms of fundamentality does not require that there be a maximally fundamental level. In a situation of “infinite descent” of fundamentality, physicalism requires only that, after a certain “depth”, everything is physical. Second, physicalism does not require that the physical fundamentalia be microscopic. Physicalism is compatible with fundamental particles the size of bowling balls and with fundamental properties that apply only to entire systems such as the whole universe. But contemporary physics seems to suggest that the fundamental physical features of our world are microscopic. If physicalism is true about the actual world, the most likely form is microphysicalism. For ease of exposition I’ll treat the two as equivalent. For more on this conception of physicalism, which makes essential use of the notion of fundamentality, cf. chapter 1 of this dissertation.

²Phenomenally conscious experiences are associated with raw experiences or sensations. Some examples include the sensation associated with a pain, the color-experience associated with seeing the sky (a “blue experience”), or the sensation one has when the back of one’s neck itches. In the words of Nagel (1974), there is something “it is like” for the subject of conscious experience. Whenever I speak of ‘consciousness’ or ‘the phenomenal’ or ‘the experiential’, I speak of phenomenal consciousness. For clarification cf. Block (1995).

then links these facts about the a priori to claims about metaphysical modality, and finishes with the startling claim that physicalism is false. (Cf. section 2 for discussion of the master argument). Many contemporary arguments against physicalism share the basic form of the master argument. These arguments include the knowledge argument of Jackson (1982, 1986), the conceivability arguments of Chalmers (1996a, 2003a, 2010), the modal arguments of Kripke (1972), and the explanatory gap argument against physicalism (Levine (1983)). One can view all these arguments as variations on one master argument: *The Master Argument Against Physicalism*.³

In this paper, my goal will be to present the master argument and investigate, more deeply than has been done before, the rationale behind it. An important motivation behind the master argument, in particular behind the key move from an epistemic to a metaphysical gap, is modal rationalism. *Modal rationalism*, in its most general form, is the view that there are important, potentially constitutive, connections between metaphysical modality and a priori justification. The master argument relies on ambitious theses in the philosophy of modality. In particular, it relies on a strong form of modal rationalism.

I have sympathies with modal rationalism. However, I deny the strong version of modal rationalism that underwrites the master argument against physicalism. I deny partly because I am a physicalist. But thankfully, I have an explanation, a diagnosis, of where strong modal rationalism goes wrong. The mistake of strong modal rationalism is a meta-semantic one. Strong modal rationalism relies on a mistaken, and overly optimistic, view of how our representations ac-

³Each argument has more to it than the master argument contains. Each argument latches onto its own particular features of the mind/body problem. But I think it helpful to consider the ways in which the arguments are similar. This kernel of similarity may tell us something central about the relationship between consciousness and the physical.

quire the meanings they do. In the end, I'll use the meta-semantic lessons drawn from the paper to offer a suggestion for how to fix the modal rationalist agenda. An important idea will be there is an important link between demonstratives and our access to and representations of conscious experience. This connection, and not a metaphysical gap stemming from the truth of dualism, is the source of the epistemic gap between the physical and conscious experience.

2.2 The Master Argument Against Physicalism

2.2.1 From Fundamentality to Modality to A Priori Implication

I take it as a datum that our world has more and less fundamental features. Quarks, leptons, and the weak nuclear force are likely fundamental. Cameras, cantaloupes, coffees, and koalas are not. The cameras, cantaloupes, coffees and koalas are the way they are in virtue of the way the quarks, leptons, and weak nuclear forces are. The non-fundamentalia are generated by, and metaphysically depend on, the fundamentalia. I use 'fundamental' as a label for the basic building blocks of the world and 'generation' as a label for the "making so" relation between those fundamental building blocks and the fundamentalia on which they metaphysically depend. (For more on the notions of fundamentality, generation, and their use in defining physicalism cf. Chapter 1.2).

Fundamentalia need not be *part* (mereological or otherwise) of the things they generate. For example, I think it likely that the world's fundamental building blocks include physical laws. These laws help to metaphysically generate the non-fundamentalia, including the cameras, cantaloupes, coffees, and koalas. But these laws are in no sense part of the cantaloupes or koalas.

The fundamental constituents of reality must, taken together, generate all of

reality. Everything is either fundamental or generated from the fundamental.⁴ There are modal constraints on the fundamental. The fundamental state of the world must metaphysically necessitate the complete state of the world. In other words, no two possible worlds can differ without differing fundamentally.

Physicalism has received a variety of (surprisingly similar) modal definitions over the years. Modal definitions appear in Lewis (1983), Jackson (1993), Chalmers (1996a), Leuenberger (2008), and Stoljar (2009b). All of these definitions boil down to some variation on the claim that no two worlds can differ without differing physically. Notice that the modal constraint on the fundamental, in combination with physicalism as a claim about the nature of the fundamental, entails the modal definitions offered by these authors. I believe that modal definitions of physicalism are actually attempts to get at the core notion: fundamentality. The use of modality instead of fundamentality is understandable. Modality is a much better understood and more respectable philosophical notion with a well-understood formal framework underpinning it.

Modality thus provides a useful tool for testing theories of the fundamental. For example, if physicalism is true, the physical state of the world necessitates the complete state of the world. If it does not, physicalism is false. If idealism is true, then the mental state of the world necessitates the complete state. And so on.

But how are we to know whether the physical state of the world necessitates the complete state? Could a world be identical to the actual world with respect to its physical fundamentalia, but differ in some other way, perhaps with respect to how the consciousness is distributed? Considerations of conceivability are one

⁴Elsewhere (chapter 1), I've called this requirement on fundamentality "the covering constraint". Schaffer (2010): 38-9 has a similar "tiling constraint".

way to get a grip on what is and is not metaphysically possible. Unfortunately, conceivability considerations seem to tell against physicalism. Situations that are physical duplicates of the actual world, but not duplicates simpliciter, are widely taken to be conceivable.⁵ Such situations include philosophical zombies and inverted qualia. (For further discussion cf. section ??: “Premise (P1): The Epistemic Gap”.)

One can also get a grip on necessitation via the relation of *a priori implication*. For any two propositions ϕ, ψ , ϕ *a priori implicates* ψ if and only if the material conditional $\phi \rightarrow \psi$ is knowable a priori. The relation can also hold between sets of propositions (defined in the obvious way), or between sentences (with meaning attached).⁶ All truths knowable a priori (such as “ $2+2 = 4$ ”) are trivially a priori implicated by any truth or set of truths, including the null set.

Let **Fund** be a proposition that completely described the layout of the world’s fundamentalia. Fund describes how all the fundamental particles, properties, and laws are distributed. If Fund completely describes the fundamental, then Fund will necessitate all truths. Here’s the additional claim: If Fund necessitates all truths, it must also a priori implicate all truths. If one knew that Fund were the case, then one could, using purely a priori reasoning, come to know each and every truth (in principle).

This claim about the a priori implications of Fund is bold, but not outlandish. The state of the cameras, cantaloupes, coffees, and koalas, is generated by and

⁵Physicalists and anti-physicalists alike seem to acknowledge the conceivability of such situations. Cf. Balog (1999), Chalmers (1996a, 2003a, 2010), Block & Stalnaker (1999), Leuenberger (2008), Hill (1997), lots of others.

⁶For any two sets of propositions Γ, Δ , Γ a priori implicates Δ iff the proposition formed by conjoining the members of Γ a priori implicates the propositions formed by conjoining the members of Δ .

metaphysically depends on the state of the fundamentalia, which is complete described in Fund. In some sense, the information about the cameras and koalas is already present in Fund.

In this paper, a priori implication will play the role I've just described. A priori implication will serve as a test for necessitation, which in turn serves as a test for fundamentality. Of course, one might wonder whether a priori implication provides a good test for necessitation in the first place. After all, a priori implication is a purely epistemic relation whereas necessitation is soundly metaphysical. Furthermore, we'll see that the master argument against physicalism relies in a crucial way on a connection between a priori implication and necessitation. One might reject *tout-court* all connections between a priori implication and necessitation and in so doing sidestep the master argument.

I do not think that this dismissive move is the way to go, but I cannot fully explain my reasons here. First, epistemic relations such as a priori implication are our main philosophical anchor in the metaphysics of fundamentality. Without epistemic considerations to guide us, we'd be lost in the search for the nature of the fundamental. Second, the connection between a priori implication and necessitation gets support from modal rationalism. I have sympathies with modal rationalism, and I hope that after reading section 5 ("Motivations for Modal Rationalism"), the reader will too. Modal rationalism is an attractive philosophical doctrine. We should preserve some form of the view if we can. Rejecting outright the master argument and its posited connections between a priori implication and necessity fails to achieve this goal. If possible, it's preferable to reply to the master argument while retaining some form of modal rationalism and avoiding the dismissive reply. In this paper, I'll try to do exactly that.

2.2.2 The Master Argument

The Master Argument is basically the following. We start from the premise that there is no a priori route from a complete description of the microphysical world to a complete description of the world simpliciter. This is so partly because there is no a priori route from the facts about microphysics to the facts about phenomenal conscious experience. The second premise is that if there is no such a priori route there is no metaphysical necessitation from the microphysical state of the world to the complete state of the world. The third premise is that if there is no such necessitation, fundamentality physicalism is false. Conclusion: fundamentality physicalism is false.

Let \mathcal{P} express the complete fundamental physical state of the world.⁷ (If fundamentality physicalism is true, \mathcal{P} expresses the complete fundamental state.) Let \mathcal{J} be an indexical truth. The indexical truth identifies a specific individual, location, and time as I/me, here, and now. \mathcal{J} turns the physical description \mathcal{P} into a centered description. The indexical truth will be discussed in section 3.5 (“Strong Modal Rationalism”), but for now I’ll ignore complications surrounding \mathcal{J} . Let \mathcal{Q} be a conjunction of all the truths about conscious phenomenal experience. \mathcal{Q}

⁷ \mathcal{P} should be understood to contain what Chalmers & Jackson (2001) call a “that’s all” clause, which effectively states that \mathcal{P} is a complete description of the fundamental state of the world. Without a “that’s all” clause, a description of the fundamental level won’t put one in a position to rule out certain possibilities for how the world is, and thus won’t permit a priori reasoning to all truths. For example, a fundamental physical description without a “that’s all” clause won’t enable one to rule out the existence of ghostly non-physical angels.

There are interesting questions about whether the “that’s all” clause expresses a fundamental feature of the world, or whether it is merely an enabling condition that allows the genuine fundamentalia to make everything the case (including the non-existence of angels). Similar questions arise about whether the “that’s all” clause should be counted as a “physical”. At the least, it seems like a topic-neutral truth whose occurrence in a description of the world’s fundamental level should not thereby jeopardize physicalism. (In the same way, the need to use mathematics and logic, which are not strictly “physical”, in a description of the world’s fundamental level does not jeopardize physicalism.) For more on this cf. Chapter 1 (“Fundamentality Physicalism”). From here, I’ll suppress issues about the “that’s all” clause. They do not alter the dialectic.

states who had what type of experience and when. ‘AP(ϕ)’ says that ϕ is knowable a priori. ‘ $\Box(\phi)$ ’ says that ϕ is metaphysically necessary. ‘ $\phi \rightarrow \psi$ ’ expresses the material conditional “If ϕ then ψ ”. With this notation, we express the master argument as follows.

(P1) not-AP($\mathcal{P}\mathcal{J} \rightarrow \mathcal{Q}$)

(P2) $\Box(\mathcal{P} \rightarrow \mathcal{Q}) \rightarrow \text{AP}(\mathcal{P}\mathcal{J} \rightarrow \mathcal{Q})$

(P3) If fundamentality physicalism is true, then $\Box(\mathcal{P} \rightarrow \mathcal{Q})$.

(C) Fundamentality physicalism is false.

The argument is valid. We investigate each premise.

2.2.3 Premise (P3): From Fundamentality to Modality

Premise (P3) is not controversial. Part of what it means to say that our world is, at the most fundamental level, purely physical, is to say that fixing how the world is at the purely physical level fixes (i.e. necessitates) how the world is simpliciter, including how the conscious states are.⁸ The fundamental state of the world metaphysically necessitates the complete state of the world, including the facts about consciousness. If \mathcal{P} fails to metaphysically necessitate the facts about consciousness, it fails to necessitate the complete state of the world. Thus \mathcal{P} fails to tell the complete story about the fundamental state of the world. There is more to the fundamental than the physical. Fundamentality physicalism is false.

2.2.4 Premise (P1): The Epistemic Gap

Premise (P1), not-AP($\mathcal{P}\mathcal{J} \rightarrow \mathcal{Q}$), says that there is no a priori route from the physical truths (supplemented with an indexical truth) to the truths about con-

⁸See Chapter 1.

consciousness. You can know all you like about the physical, but this won't tell you what it's like to see red, or whether some creature is experiencing blue, green, pain, or nothing at all.

Many sources support (P1). I'll discuss three such sources here: (i) the explanatory gap between the physical and consciousness (ii) the knowledge argument thought experiment of Jackson (1982, 1986) and (iii) the conceivability of inverted qualia and philosophical zombies.

Consider the famous explanatory gap between the physical truths and the truths about conscious experience (cf. Levine (1983), Chalmers (2003a)). If there is an explanatory gap between the physical and consciousness then there is, *ipso facto*, an a priori explanatory gap. The famous Mary thought experiment of Jackson (1982, 1986) provides additional support. Jackson asks us to envisage Mary, "a brilliant scientist who is... forced to investigate the world from a black and white room via a black and white television monitor. She... acquires... all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on." Jackson asks us to consider whether Mary learns anything new when she is released from the black and white room. Jackson maintains that it just seems obvious that she will "learn something about the world and our visual experience of it." Mary knows all the physical truths in \mathcal{P} . But she can't come to know all the truths, primarily because she can't come to know truths about conscious experience. If Mary can't figure it out on the basis of \mathcal{P} , no one can. There's no epistemic route from \mathcal{P} to consciousness. If there's no epistemic route, there's no a priori route. Therefore: (P1) is true.

Conceivability arguments also support (P1). I'll consider two: the inverted

qualia argument and the zombie argument. Both appear in Chalmers (1996a). The inverted qualia argument starts from the claim that we can conceive of a physical duplicate of the actual world in which the conscious experiences are switched. When you have a blue experience, your inverted twin has a red experience. When you have a painful sensation, your twin has a tickle. Et cetera. The zombie argument asks us to conceive of a world in which your twin is a philosophical zombie, and has no conscious experiences whatsoever. These zombies walk the walk, and talk the talk, and even utter sounds like, “My, that’s a beautiful shade of green.” But these zombies have no experience on the inside. The next stage of the argument infers from the conceivability of these situations to their metaphysical possibility. If they are metaphysically possible, physicalism is false. (The move from conceivability to possibility in these arguments is the analogue of the master argument’s premise (P2)).

Chalmers (and most others) are clear that conceivability is an epistemic notion. It has to do with what one can coherently imagine. The “coherently” bit is important, as is the idealization of conceivability. One might think that one could, in the relevant sense, conceive that $2 + 2 = 5$, or that there are finitely many prime numbers, or even that a microphysical duplicate of the actual world lacked chairs. But once the notion of conceivability has been idealized to make the conceivers highly intelligent, maximally consistent, and hyper rational, these scenarios will not be conceivable in the relevant sense.

For Chalmers in particular, conceivability is tied to the a priori.⁹ ϕ is conceivable if and only if it’s not a priori that not- ϕ .¹⁰ On this construal, to say that the

⁹Chalmers (1996a, 2002).

¹⁰Chalmers (2002) calls this ‘ideal negative conceivability’.

zombie world is conceivable is to say that a proposition Z which conjoins all the physical truths and denies the existence of conscious experience is conceivable, i.e. can't be ruled out a priori.¹¹

We can express the proposition Z as " \mathcal{P} and no- Q ", where \mathcal{P} describes the fundamental microphysical layout of the world, and no- Q says that there is no conscious experience. Thus to say that zombies are conceivable, is, for Chalmers, to say that " \mathcal{P} and no- Q " can't be ruled out a priori. If this is so, then there can be no a priori implication between \mathcal{P} and no- Q . If \mathcal{P} did a priori implicate no- Q , then " \mathcal{P} and no- Q " could be ruled out a priori. On this construal of conceivability, the conceivability of zombies deductively entails premise (P1): not-AP($\mathcal{P} \mathcal{J} \rightarrow Q$) (suppressing issues surrounding \mathcal{J}).¹² The conceivability of inverted qualia has the same ramification regarding a priori implication.

2.2.5 Premise (P2): From an epistemic to a metaphysical gap

The second premise (P2) is more controversial. It endorses a connection between a priori justification and metaphysical modality. More precisely, it claims that if all the facts cannot be known a priori on the basis of the physical facts, the physical facts do not necessitate all the facts. (P2) is an instance of a more general schema. The generalized schema says that if a set of truths necessitates all truths, and meets certain further criteria (to be discussed later), then that set of truths also a priori implicates all truths. The schema says that if physicalism were true, then an idealized reasoner who knew everything there was to know about the

¹¹In the terminology of Chalmers (2002), this notion corresponds to what he calls "ideal negative conceivability".

¹²Strictly speaking, the argument claims that since $\mathcal{P} \mathcal{J}$ and no- Q is conceivable, it is possible, and physicalism is false. If $\mathcal{P} \mathcal{J}$ and no- Q is conceivable, then there is no a priori implication from $\mathcal{P} \mathcal{J}$ to no- Q .

fundamental particles, properties, and laws of our world would be in a position to know everything. She could figure out all the facts about cameras, cantaloupes, coffees, and koalas.

(P2) embodies a litmus test for fundamentality. Consider some alleged story, X-ism, about the fundamental nature of the world: physicalism, dualism, idealism, deism, tomato-ism whatever. If X-ism is true, then the X-ical truths will tell the complete story of the fundamental structure of the world. The X-ical truths will necessitate all truths. If they necessitate all truths, they should a priori implicate all truths (or so the schema maintains). The Master Argument maintains that since the physical truths fail to a priori implicate all truths, the physical truths do not provide a complete account of the fundamental. Physicalism is false.

This litmus test for fundamentality relies on a particular connection between necessitation and a priori implication. Such connections tend to be endorsed, in some form or other, by modal rationalists. This brings forth the obvious question: “Why should metaphysical claims about necessitation and modality follow from epistemic claims about what can be known a priori on the basis of something else?” Answering this question requires exploration of deep waters in the philosophy of modality and of meaning. It is into those waters that we now voyage. We will spend most of the paper there.

2.3 Supplementation

2.3.1 Covering Your Bases

Let a *fact* be a true Fregean proposition. Fregean propositions are structured composites of concepts.¹³ The truth conditions of the structured composite are

¹³Throughout this paper I use a Fregean approach to propositions, according to which propositions are structured composites of concepts. This contrasts with a Russellian approach, according

determined by the references of the concepts contained in the proposition and the manner in which they are composed. For example, the proposition CRICKET IS A GAME contains the concepts CRICKET, IS, A, and GAME. (I use small capitals to denote items at the conceptual level: concepts or propositions. For example, CRICKET denotes the cricket concept.) Facts can enter into necessitation relations. *A necessitates B* if and only if the material conditional $A \rightarrow B$ is necessarily true. For any propositions ϕ, ψ , we write $\Box(A \rightarrow B)$ for “A necessitates B”. In the parlance of possible worlds, ϕ necessitates ψ if and only if ψ is true in every metaphysically possible world in which ϕ is true. For example, CRICKET IS A GAME necessitates SOMETHING IS A GAME. SOME WATER IS WET necessitates SOMETHING IS WET.

Facts also enter into relations of a priori implication. (Reminder: for any two propositions ϕ, ψ , ϕ *a priori implicates* ψ if and only if the material conditional $\phi \rightarrow \psi$ is knowable a priori.) CRICKET IS A GAME a priori implicates SOMETHING IS A GAME. One proposition can necessitate a second without a priori implicating it. For example, SOME WATER IS WET necessitates SOME H₂O IS WET.

to which propositions contain objects or properties themselves. This move is controversial, but I don't think much turns on this choice in the end. Our target phenomenon, the relations between necessity and the a priori, provide the main reason for opting for a Fregean approach. Russellian propositions are simply not well suited to such an investigation. To start, I'm unsure how to describe the phenomenon of a priori knowability on a Russellian approach. For the Russellian, the sentences 'Hesperus is Hesperus' and 'Hesperus is Phosphorous' express the same proposition: $\langle \odot = \odot \rangle$ (where \odot is Venus itself (i.e. Hesperus, i.e. Phosphorous)). But one can know a priori that Hesperus is Hesperus, whereas that Hesperus is Phosphorous is a paradigm of knowability only a posteriori. (Perhaps the Russellian will deny this?) On the Russellian approach, it looks like propositions are not the bearers of a priori / a posteriori knowability. (At the least, they're not the complete bearers). To account for the differences between the a priori knowability of "Hesperus is Hesperus" and "Hesperus is Phosphorous", the Russellian must appeal to additional tools. Following the suggestions of Salmon (1986), one might appeal to Russellian-propositions-under-guises. Proposition P could be knowable under one guise but not under another. I hypothesize that in the end, the Russellian account, appropriately supplemented to deal with the a priori (perhaps via guises), will behave much like a Fregean account of propositions. I don't care much which of the available entities we dub 'propositions'.

But SOME WATER IS WET does not a priori implicate SOME H₂O IS WET, because the relationship water and H₂O is knowable only a posteriori.

With these definitions in tow we define three more notions: (i) necessitation base (ii) a priori implication base, and (iii) fundamentality base. A set of propositions S is a *necessitation base* if and only if the conjunction of all propositions in S necessitates every true proposition. A necessitation base fixes all the facts. It leaves the truth-value of no proposition undecided. Some examples of candidate necessitation bases include the set of all microphysical truths (if microphysicalism is true), the set of all microphysical truths plus truths about phenomenal experience (if some type of dualism is true), and the set of all truths believed by God (if God is omniscient). Trivially, the set of all truths (if there is such a thing) is a necessitation base, because every truth necessitates itself.

A set of truths S is an *a priori implication base* if and only if the conjunction of all truths in S a priori implicates every truth. A useful heuristic for thinking about a priori implication bases goes as follows. If one knew all the truths in an a priori implication base, one could, in principle, use purely a priori reasoning to come to know everything.¹⁴ Examples of candidate a priori implication bases include the set of microphysical truths (if a priori physicalism is true) and the set of microphysical plus phenomenal truths.

A set of truths is a *fundamentality base* if it fully describes the fundamental state of the world. If physicalism is true, then some set of physical truths provides a fundamentality base. If dualism is true, then some conjunction of physical and

¹⁴This heuristic may be only a heuristic, because of problems related to unknowable truths. If a truth T is unknowable, then no amount of a priori reasoning will allow a reasoner to know T. However, T's unknowability does not conflict with the ability of a base of propositions B to a priori implicate T. Even if T is unknowable, the conditional B \rightarrow T might still be knowable a priori. Cf. Chalmers (2012), chapter ??.

mental truths provides a fundamentality base. Any fundamentality base must also be a necessitation base. This claim is partly constitutive of the notion of the fundamentality.

2.3.2 The Interest of the Base

Why should one be interested in necessitation, or a priori implication, bases in the first place? In general, necessitation is a useful test for constitutive connections. If x is constitutively F , then, necessarily, x is F (the converse implication does not always hold). Thus necessitation serves as a test for constitutivity.

One major task of both metaphysics and science is to determine the fundamental constituents of reality, the Aristotelian substances.¹⁵ One already mentioned restriction on the fundamental constituents of reality is that they, together, generate or give rise to the rest of reality. This entails that the fundamental constituents of reality must necessitate all truths. Any fundamentality base must be a necessitation base.

In the physicalism-dualism debate, necessitation (or lack thereof) serves as an indicator (or contra-indicator) of fundamentality. In the first paragraph of this paper, I defined fundamentality physicalism as the thesis that our world is, at the fundamental level, purely physical. This means that the fundamentality base for our world is a set of purely physical facts. These propositions mention quarks, gravity, and space-time, but not government, grass, or green sensations. This set

¹⁵In *Metaphysics*, Aristotle writes that "Substance is the subject of our inquiry; for the principles and the causes we are seeking are those of substances. For if the universe is of the nature of a whole, substance is its first part..." (1984: 1688; *Meta.1069a18*–*20*). Gill (1989) summarizes Aristotle's test for substance-hood as follows: "the main criterion [for selecting the primary substances] is ontological priority. An entity is ontologically primary if other things depend for its existence on it, while it does not depend in a comparable way on them." I described the fundamental entities as those on which everything else metaphysically depends, and which depend on nothing further.

of purely physical facts, the physicalist fundamentality base, must necessitate all facts.

Fundamentality dualists argue that purely physical facts (pick any ones you like) don't necessitate all facts. In particular, they don't necessitate all the mental facts. Thus the purely physical facts are not a fundamentality base, and physicalism is false. We thus see that in the physicalism/dualism debate, a main point of contention is whether some appropriate set of physical facts is a necessitation base.

In sum, the notion of a necessitation base is a vital tool for evaluating claims about the fundamental structure of the world. Fundamentality physicalism is a claim about that structure, and determining whether or not the physical can provide a necessitation base is a vital step in the investigation of physicalism.

The interest of an a priori implication base is less obvious. Modal rationalists will tend to claim that a priori implication is a guide to necessitation, in the same way that necessitation is a guide to fundamentality. Thus some of the interest in necessitation bases will transfer to a priori implication bases.

The search for an a priori implication base falls within the philosophy of scrutability. A set of truths is *scrutable* from a second set when the second set bears some relation of epistemic entailment to the first. A priori implication provides one type of epistemic entailment. A posteriori implication, logical entailment, and derivability (of various types) provide others. Scrutability bases provide a guide to the conceptual or epistemic structure of the truths about a given domain (the largest domain being the domain of all truths). Here conceptual or epistemological structure is contrasted with metaphysical structure, and implication is the epistemological analogue of metaphysical necessitation. A priori implication

bases may also offer a guide to a certain type of conceptual fundamentality.¹⁶ (An excellent discussion of scrutability in general, as well as the applications and interest of scrutability bases, can be found in Chalmers (2012)).

Foundationalist projects in mathematics start from a few basic principles (called ‘axioms’), introduce a system of proof, and attempt to derive (within that proof-system) all the other truths about the domain. The search for an a priori implication base can be seen as a similar task, with a priori implication playing the role of the method of “proof”. Here, we search for an a priori implication base for all truths, not just mathematical truths. But the notion of an a priori implication base can easily be restricted to any domain of truths.

Rudolf Carnap (1929)’s *Der logische Aufbau der Welt* (“The Logical Structure of the World”) attempts to develop a minimal epistemological base for all truths. For Carnap, the implication relation is logical implication, and the base consists of truths expressible using logical expressions plus a primitive 2-place relation of phenomenal similarity. This is a small base indeed; it would be an incredible discovery if all the truths about the world were logically derivable from sentences expressed using only logic plus phenomenal similarity.

The search for an a priori implication base lies in the same vein as Carnap’s project. One replaces logical implication with a priori implication as the method of getting from the base truths to other truths. But, like Carnap’s base, an a priori implication base purports to be an epistemological source for all truths (not just, e.g., mathematical truths). I find the following question fascinating: “What is the minimal set of truths from which all truths are a priori implicated?” Questions about the language or vocabulary that will be used to express the base truths are

¹⁶For more on conceptual fundamentality, and conceptual grounding, cf. Chalmers (2012), esp. Excursus 16 (pages numbers here).

equally compelling. Of course, such epistemological bases may not be unique. But the variety of such bases may yet teach us about the conceptual structure of the world.

2.3.3 The Supplementation Question

Once we've introduced the two types of bases we can ask some obvious questions. What is the smallest possible base (of any type)? How should we measure the size of a base? Counting the number of propositions will not work. One can always use conjunction to create one proposition out of two. In this paper, our primary interest in a priori implication bases will be in their role as guides to necessitation, which in turn are guides to fundamentality. Consider the supplementation question:

The Supplementation Question: "What must be done to a necessitation base to create an a priori implication base?"

There's an important link between modal rationalism and the supplementation question. Modal rationalists believe that there are important, potentially constitutive, connections between metaphysical modality and a priori justification. Modal rationalists are likely to think that there are interesting answers to the supplementation question. Anti-rationalists, on the other hand, will think that there are no interesting answers to the supplementation question.

There is much more to modal rationalism, in any form, than its answer to the supplementation question. In this paper, we are interested in modal rationalism because of its role in the physicalism-dualism debate as the chief sponsor of premise (P2) of the master argument. As a result, we will use the supplementation question to measure the strength of a rationalist view. Stronger forms of

rationalism will claim that a necessitation base needs less supplementation to become an a priori implication base.

2.3.4 Naive Modal Rationalism

According to *naive modal rationalism*, every necessitation base is already an a priori implication base. No supplementation is required. This view, however, cannot be sustained. The basic reason for this is that not all necessitation is a priori. Truths containing HESPERUS necessitate but don't a priori implicate truths containing PHOSPHOROUS. It does not immediately follow that naive modal rationalism is false. All we've established so far is that there are some propositions A, B such that A necessitates B without a priori implicating B. It may yet be that, for the special classes of propositions that form necessitation bases, the proposition formed by conjoining all the members of the necessitation base does a priori implicate all the truths it necessitates. If so, naive modal rationalism would be vindicated despite the existence of a posteriori necessitation.

Unfortunately, there are trivial necessitation bases that are epistemically opaque, in the sense of a prior implicating very little at all. As a result naive modal rationalism is false. The proposition EVERYTHING IS JUST AS IT ACTUALLY IS is a necessitation base (henceforth, we call this proposition '@').¹⁷ @ trivially necessitates every necessary truth (because any proposition necessitates the necessary truths). So it remains to be shown only that @ necessitates the contingent truths. @ necessitates every contingent truth iff, for every contingent truth C, every possible world at which @ is true is a world at which C is true. Assume that @ is true at arbitrary world w. Suppose, for reductio, that C is false at w. Because C is false at w and true at the actual world, at w everything is not just as it actually is.

¹⁷I owe this example to Schwarz (2007).

@ is false at w , which contradicts our assumption. We reject the supposition that C is false at w . Therefore, @ necessitates all contingent truths in addition to all necessary truths. @ is a trivial necessitation base that a priori implicates little to nothing. Not all necessitation bases are a priori implication bases. Naive modal rationalism is false.

Indexical truths provide another reason to reject naive modal rationalism. A set of propositions that describes the world completely from a 3rd person or objective standpoint but does not include any indexical truths can succeed in necessitating all truths but fail to a priori implicate them all. For example, such a set of truths might explicitly declare that Gabriel Rabin has a sweet moustache, and thereby necessitate that I have a sweet moustache. But such a set of truths will fail to a priori implicate that I have a sweet moustache (here I assume that I AM GABRIEL RABIN is not knowable a priori).

The problem with indexicals runs beyond failure to a priori implicate truths containing indexicals. Suppose one has a necessitation base that describes the entire universe. One can tell immediately that the universe contains two planets, similar in every way compatible with the following facts. On the first planet the clear wet liquid that flows in the lakes, rivers, and streams is the chemical compound H_2O . On the second planet the clear wet liquid that flows in the lakes, rivers, and streams is a superficially similar substance with a different underlying chemical make-up: XYZ. One cannot tell a priori, from this description, which planet contains water and which twin-water. In order to determine whether one's concept WATER refers to H_2O or XYZ, one first needs to know which planet one inhabits. If one inhabits the H_2O planet, then one's WATER representations refer to H_2O . If one inhabits the XYZ planet, one's WATER representations refer to

XYZ (more accurately, one's seemingly WATER representations are actually TWIN-WATER representations). One needs, in the a priori implication base, a truth that locates oneself on one planet or the other.¹⁸ The basic point of the example is that indexical truths are required for the a priori implication even of non-indexical truths.

In sum, there are many reasons to reject the naive modal rationalist answer to the supplementation question. Metaphysical modality and the a priori do not have a tight enough link to support the naive reply. However, the failures of naive modal rationalism offer lessons for how to improve one's answer to the supplementation question. A form of modal rationalism that incorporates these lessons may yet succeed.

2.3.5 Strong Modal Rationalism

The lesson of @ is that some propositions are epistemically opaque. They necessitate a lot but a priori implicate little. The lesson of I HAVE A SWEET MOUSTACHE and WATER IS H₂O is that any a priori implication base must include some indexical truths. Strong modal rationalism heeds these two lessons in its answer to the supplementation question. Strong modal rationalists think that, in order to become an a priori implication base, a necessitation base must (i) be expressed in a canonical semantically neutral vocabulary and (ii) be supplemented with an indexical truth J.

The strong modal rationalist answer to the supplementation question yields the following principle:

- **(Strong Modal Rationalist Link):** A canonically expressed semantically

¹⁸The locating must be done in a first-person way. The proposition GABRIEL IS ON THE H₂O PLANET will not suffice.

neutral necessitation base, supplemented with an indexical truth, a priori implicates every proposition it necessitates.

An *indexical truth* is a truth that selects a privileged location, time, and person as here, now, and I, respectively. A necessitation base is semantically neutral when it is expressed in a canonical vocabulary: a vocabulary that is semantically neutral. Semantic neutrality is first and foremost a property of concepts (or terms). Semantic neutrality is a difficult notion to pin down, but the idea is clear enough. Intuitively, a concept is semantically neutral iff it is not “twin-earthable” (in the sense of Putnam (1975)).¹⁹ Famously, ‘water’ is twin-earthable. You can refer to H₂O while your twin-earth duplicate refers to XYZ. Some paradigms of twin-earthable terms include ‘water’, ‘gold’, ‘Hesperus’, ‘actual’, and almost any proper name or natural kind term. Some plausibly semantically neutral terms include ‘friend’, ‘bachelor’, ‘justice’, ‘morally right’, ‘cause’, and the logical connectives. (A more detailed discussion of semantic neutrality occurs in section 2.4.2: “Semantic Neutrality”).

A proposition is semantically neutral if all the concepts it contains are semantically neutral. A set of propositions is semantically neutral if all the propositions it contains are semantically neutral. Semantically non-neutral propositions tend to necessitate much that they do not a priori implicate. Non-neutral representations are epistemically opaque. This is why lack of neutrality impedes a necessitation base’s aspirations for a career as an a priori implication base.

Let’s sum up. The supplementation question asks, “What must one do to a necessitation base to create an a priori implication base?” The strong modal

¹⁹For our purposes here, terms for which one can run a twin-earth thought experiment in the vein of Burge (1979) do not count as “twin-earthable”. Any term is twin-earthable in that sense.

rationalists answer, “If the necessitation base is canonically expressed, add an indexical truth.” (In section ??: “Semantic Neutrality and the Physical Truths”, I address questions about how one might “neutralize” a non-neutral base.) Importantly, if a necessitation base is (a) canonically expressed and (b) contains an indexical truth, then strong modal rationalism says that the necessitation base is also an a priori implication base. This is the key claim that leads to the tension between strong modal rationalism and physicalism.

2.4 Return to the Master Argument

In this section I precisify the master argument and explain how it relies on the strong modal rationalist’s answer to the supplementation question.

2.4.1 The Strong Modal Rationalist’s Test for Fundamentality

The strong modal rationalist’s answer to the supplementation question, embodied in the strong modal rationalist link, offers a test for necessitation bases. The test goes like this. First step: check whether the alleged necessitation base, a set of propositions, is expressed in a canonical semantically neutral vocabulary. If it is not, express it in a semantically neutral way. Once the base is semantically neutral, move to the next step. Second step: check whether the base contains an indexical truth \mathcal{J} . If it does not, add one. Third step: check whether the result of the first two steps is an a priori implication base. If it is not, the set of truths from which we began is not a necessitation base. (This test will not always yield a verdict. If the set of truths is not semantically neutral, and can’t be molded into a semantically neutral form, then the strong modal rationalist test yields no verdict on whether they form a necessitation base.)

The master argument applies the strong modal rationalist’s test to argue that

fundamentality physicalism is false. The basic argument goes as follows. If physicalism is true, then some set of physical truths \mathcal{P} is a necessitation base. Such a set of truths is either semantically neutral or can be canonically expressed in a neutral fashion. An indexical truth can be added. But the result is not an a priori implication base. Why not? Because there is no a priori route from the physical truths to the truths about phenomenal conscious experience. The purely physical truths won't enable one to determine whether some thinker is experiencing red, blue, or nothing at all. Therefore, the set of physical truths is not a necessitation base. Fundamentality physicalism is false.

For clarity's sake, I repeat the master argument from section 2.1:

(P1) not-AP(\mathcal{P} \rightarrow \mathcal{Q})

(P2) $\Box(\mathcal{P} \rightarrow \mathcal{Q}) \rightarrow \text{AP}(\mathcal{P} \rightarrow \mathcal{Q})$

(P3) If fundamentality physicalism is true, then $\Box(\mathcal{P} \rightarrow \mathcal{Q})$.

(C) Fundamentality physicalism is false.

These premises were already discussed in section 2. We won't question (P1) by denying the epistemic gap between the physical and conscious experience. Nor will we question (P3), which is partly constitutive of physicalism. (P2) is the main suspect. Two issues remain to be discussed. First, is the strong modal rationalist's answer to the supplementation question correct? I.e. is the litmus test embodied in (P2) a good one? In section ?? ("Against Strong Modal Rationalism"), I'll argue that it is not.

Second, is \mathcal{P} semantically neutral or neutralizable? If it is not, then the strong modal rationalist's test does not apply. The next section ("Canonical Expression

of the Base: Semantic Neutrality”) discusses semantic neutrality and \mathcal{P} ’s status with respect to it.

2.4.2 Canonical Expression of the Base: Semantic Neutrality

Modal rationalists, as I have characterized them, want to use a priori implication (or lack thereof) as a guide to necessitation and necessitation as a guide to fundamentality. Anyone who endorses such a strategy should insist that the classes of truths tested be expressed in a canonical fashion. Allowing any mode of expression immediately robs the project of philosophical interest.

As an example, consider the proposition EVERYTHING FUNDAMENTAL IS JUST AS IT ACTUALLY IS. This proposition is a necessitation base (assuming the fundamentality-modality link discussed in section 2.2.1²⁰). But this proposition a priori implicates almost nothing. If we allows propositions such as these, the project of finding an a priori implication oriented test for necessitation is a non-starter. Terms like ‘actual’ make it too easy to necessitate a lot while a priori implicating very little.

Furthermore, despite the the proposition EVERYTHING FUNDAMENTAL IS JUST AS IT ACTUALLY IS’s credentials as a necessitation base, it tells us almost nothing about the nature of the fundamental. It provides no help adjudicating the dispute between the fundamentality physicalist and the fundamentality dualist.

‘Actually’ is the problematic term. ‘Actually’ is epistemically opaque. It allows reference to things to which it’s not clear reference is being made. For example, ‘everything fundamental is just as it actually is’ makes reference to the complete fundamental layer, and in some sense describes that layer. But from an epistemic standpoint, it reveals nothing about what the fundamental is like.

²⁰Premise (P3) of the master argument also requires the fundamentality-modality link. Cf. also “The Modal Covering Constraint”, in section [\[\[NOTE: Fix this citation\]\]](#) ?? (“The Covering Constraint”) of Chapter 1.

The problem of opaque vocabulary occurs at both stages of the move from a priori implication to fundamentality. It immediately makes trouble for the attempt to use a priori implication as a test for necessitation. But it also makes trouble for the attempt to use necessitation as a guide to fundamentality. For both these reasons, we should require that the relevant descriptions of the fundamental layer be expressed in a canonical, epistemically transparent, fashion. No ‘actually’s, no rigidifying operators, no ‘dthat’s (cf. Kaplan (1989)). However, this much leaves much open in terms of what transparent canonical expression requires.

Strong modal rationalists have their favored form of canonical expression: semantic neutrality. The Strong Modal Rationalist test for fundamentality applies only if the prospective necessitation base (an alleged description of the world’s fundamental level) is expressed in a canonical semantically neutral vocabulary. I will not take the time here to fully explore all the relevant issues about semantic neutrality. I’ll briefly explicate the notion to give the reader the flavor of semantic neutrality.²¹ Then I’ll explain the virtues the rationalist sees in expressing a candidate necessitation base in a semantically neutral way.

Semantic neutrality is a special epistemic property that a term or concept can have. A term or concept is semantically neutral if it has a particular link between the a priori connections of the term/concept and the intension (function across possible worlds) of the term/concept. Many terms and concepts yield a priori truths about their extension. For example, ‘bachelor’ yields the a priori truths “all bachelors are male” and “all bachelors are unmarried”. (For the sake of the example, I’ll assume that the bachelors are all and only the unmarried males).

²¹Fuller accounts of semantic neutrality can be found in Chalmers (2006b) and Jackson (1998).

Let's call this set of a priori truths the *a priori conceptual role*, or just *ap-role*, of the term/concept. The ap-role yields an intension: a function across worlds. The ap-role-intension of 'bachelor' picks out the unmarried males across all possible worlds. Importantly, so does the familiar intension of 'bachelor' from possible worlds semantics. A term is semantically neutral when its ap-role-intension matches its normal intension. The term 'bachelor' and the concept BACHELOR are semantically neutral.

Many terms are not semantically neutral. For example, suppose that the a priori conceptual role of 'Hesperus' is "the brightest celestial body visible in the evening sky". The ap-role-intension of 'Hesperus' picks out, at each world, the brightest celestial body visible in the evening sky. At some worlds, this will be Hesperus, i.e. Venus. At some worlds it won't be. The ap-role-intension of 'Hesperus' does not match the normal intension, which picks out Venus at every possible world. Thus 'Hesperus' is not semantically neutral. To be clear: the a priori conceptual role of a term generates one intension. The term itself has another. The term is semantically neutral if and only if the two intensions are the same function.²² Names and natural kinds terms tend not to be semantically neutral.²³ For a semantically neutral term, the a priori role intension matches the custom-

²²More precisely, the a priori conceptual role determines a function across centered worlds. But the basic idea is the same.

²³Semantic non-neutrality is not equivalent to what Kripke (1972) calls 'rigid designation'. A term is a rigid designator if it picks out the same individual at every world. First, some semantically neutral terms are rigid designators. 'The number one' and 'the least prime number greater than 10' provide two examples. Second, the notion of rigid designation does not apply to non-singular terms such as 'water' and 'gold', whereas semantic neutrality applies to all terms and concepts. However, philosophers sometimes speak of mass and count nouns such as 'water' and 'gold' as rigid designators. They are easily understood when they speak in this technically incorrect fashion. I hypothesize that when philosophers speak of 'water' and 'gold' as "rigid designators", they are using "rigid designation" as a label for the phenomenon of semantic non-neutrality.

ary intension. This allows a priori access to the modal profile of the term - what it refers to at each possible world. This makes semantically neutral terms and concepts epistemically transparent in a way that non-neutral terms and concepts are not.

Semantic neutrality accomplishes the goals of a transparent canonical vocabulary. It prevents trivial necessitation bases like EVERYTHING FUNDAMENTAL IS JUST AS IT ACTUALLY IS by banning the term 'actually' as semantically non-neutral. Second, it appears to accomplish the goal of allowing necessitation to serve as a test for fundamentality. Semantically neutral necessitation bases do a better job allowing one to read off the fundamentalia from the propositions in the base.

Let me expand on this second point. We've already seen that EVERYTHING FUNDAMENTAL IS JUST AS IT ACTUALLY IS succeeds as a necessitation base but completely fails to aid inquiry into the nature of the fundamental. This might seem like an abstruse or corner case. But the lesson can be extended to situations more directly germane to the issue of physicalism vs. dualism. A physicalistically respectable description can necessitate all truths, giving the impression that fundamentality physicalism passes the necessitation test, despite the fact that fundamentality dualism is true. This can happen when the physical description is expressed in semantically non-neutral vocabulary that makes illicit and secret reference to non-physical extras.

Let me sketch an example of such a situation. Suppose that electrons are not fundamental particles. The properties of electrons are grounded in a deeper more fundamental level, consisting of an energetic fluid called 'juice'. There is two-way necessitation between the state of the juice and the state of the electron metaphys-

ically generated by that juice. (Such two-way necessitation may be rare, but it's certainly possible). The state of the world's electrons necessitates the state of the juice and vice versa. Suppose also that 'electron' designates only negatively charged particles that behave like electrons and are metaphysically generated by juice.

Under these assumptions, 'electron' is not semantically neutral. The a priori role intension of 'electron' picks out negatively charged particles at every possible world. But the intension of electron picks out only the electron-ish negatively charged particles generated by juice. Comparison with 'water', a paradigm of semantic non-neutrality, helps illustrate the lack of semantic neutrality of 'electron' in the envisioned scenario. 'Water' designates only H₂O at each possible world. At some world, a non-water substance (such as Putnam (1975)'s XYZ) could act like water but not be water, because it failed to be H₂O. Similarly, at each world 'electron' designates only the juice-generated negatively charged particles. At some world, a non-electron particle might act like an electron but not be an electron, because it failed to be generated from juice.

Given these assumptions, it's possible for a set of truths that does not mention juice, but does mention electrons, to be a necessitation base. This base could be expressed in the austere vocabulary of physics, using terms like 'electron', 'gravity', and 'strong nuclear force', supplemented with some mathematics. Such a situation will give the impression that fundamentality physicalism is true. But for all we've said so far, juice could be a Cartesian mind-substance! If so, fundamentality physicalism is clearly false, despite the a physical set of truths success as a necessitation base. Even if juice is a perfectly physical substance, and the envisioned scenario does not challenge fundamentality physicalism, the physicalist

should still want to include juice on the roster of *fundamentalia*. The success of the necessitation base above, which mentions electrons but not juice, masks the vital role juice plays as one of the world's most basic metaphysical building blocks.

One important lesson is that epistemically opaque vocabulary facilitates hidden reference to extras (physical or non-physical). If a description of the fundamental layer contains opaque (perhaps semantically non-neutral) terms, one can't read off from that description the nature of the fundamental, and thereby decide the issue of fundamentality physicalism vs. dualism. In my opinion, this is the most important reason for the rationalist to insist on canonical expression of the necessitation base. A non-canonically expressed necessitation base can fail to vindicate fundamentality physicalism.

All parties to the debate should insist that a candidate necessitation base, or description of the fundamental layer, be expressed canonically, in order to prevent illicit reference to extras, such as Cartesian mind-substances or underlying physical juices. One could hold a variety of views about the correct form of canonical expression. Modal rationalists choose semantic neutrality.

I stop to consider a worry that may have occurred to the reader. I've complained that a non-canonically expressed necessitation base fails to provide evidence of fundamentality. But necessitation of all truths by the physical truths is, at best, a necessary condition for fundamentality physicalism, not a sufficient condition. If necessitation of all truths by physical truths is not a sufficient condition for physicalism, then we should not be concerned that a physical necessitation base could succeed while fundamentality physicalism remained false. The need for canonical expression is moot.

I offer two replies to the worry. First, much of the debate surrounding the master argument operates not with fundamentality physicalism as its operative conception of physicalism, but instead with modal physicalism. *Modal physicalism* says, more or less, that the physical truths necessitate all the truths.²⁴ Necessitation of all truths by the physical truths is thus sufficient, and not just necessary, for modal physicalism (in contrast with fundamentality physicalism). When operating with modal physicalism, the need for canonical expression of the necessitation base is especially pressing. Second, we can do better than using the necessitation test as a mere check for fundamentality. If the base is canonically expressed, then success in necessitating all truths permits the base to be used as a guide to the nature of the fundamental and as strong evidence for a given theory of the fundamental.

2.4.3 Semantic Neutrality and the Master Argument

The modal rationalist's test for necessitation, embodied in premise (P2) of the mastery argument, requires that the set of physical truths \mathcal{P} be either semantically neutral or formulable using semantically neutral vocabulary. One could resist the master argument by rejecting that it is possible to express the physical truths in a semantically neutral way. One might reject the notion of semantic neutrality altogether.²⁵ Or one might accept semantic neutrality and maintain that the physical state of the world cannot be expressed in such a vocabulary. For now, I wish to put both moves aside. One of my goals is to find a form of physicalism, and a reply to the master argument, that respects and meshes with rationalist intuitions.

²⁴For more on modal physicalism, cf. Chalmers (1996a, 2010), Jackson (1993), Stoljar (2009b,a), Leuenberger (2008), Lewis (1983), and Chapter 1 of this document.

²⁵Even if one rejects the notion of semantic neutrality, the demand for canonical expression remains if the necessitation test is to serve as evidence of fundamentality.

Rejecting semantic neutrality, an important cog in the philosophical system of existing modal rationalists (including Chalmers (1996a, 2010), Jackson (1998), and Chalmers & Jackson (2001)) is a poor way to meet this goal.

From here, we assume that the physical truths can be expressed in a semantically neutral fashion, either because they are semantically neutral from the start or can be re-expressed using semantically neutral vocabulary.

2.5 Motivations for Modal Rationalism

2.5.1 Overview

According to modal rationalism, there are important, potentially constitutive, connections between metaphysical modality and a priori justification. In this section, I wish to address the following question: “Why believe strong modal rationalism, or any form of modal rationalism, for that matter?” Answering this question should both help the reader better understand what modal rationalism is and why one should believe it. I’ll canvas five motivations behind modal rationalism (MR):

1. MR has a long and classic tradition.
2. MR sanctions a ban on brute necessity.
3. MR supports and gets support from a connection between conceivability and possibility.
4. MR vindicates the factorization view of the necessary a posteriori.
5. The view from above supports MR.

2.5.2 Some Ideas Behind Modal Rationalism

First, modal rationalism is a classic position. A whole host of philosophical giants, including Kant, Leibniz, and Descartes, believed in connections between metaphysical necessity and the a priori. Of course, before Kripke (1972), many tended to conflate the notions of metaphysical necessity and a priori knowability. Kripke taught us to be more careful. But we should not throw out the baby with the bathwater by severing all such connections. (Wright (draft): 1-2 seems to agree.)

Second, the modal facts about what truths are contingent and necessary are not brute. If some truth T is necessary rather than contingent, there is an explanation for this fact. It can't be that the space of all possible worlds just happened to be such that T is true in all of them. [Quote Trogdon+Levine here?]. That's not how possible worlds work. Schwarz (2007) offers the following consideration for why there can't be brute necessities (which he calls "strong necessities"):

From a formal point of view, it is easy to see how there could be strong necessities. We only need to decouple the space of possible worlds from any links to epistemic or conceptual matters. Suppose, for instance, that there are only 7 possible worlds, all of which contain penguins. Then the existence of penguins is a strong necessity: no ordinary investigations in the actual world, and no amount of conceptual analysis, could reveal to us that penguins necessarily exist, and yet they do. The main problem with this view is that it ignores the constraints that come with the name "possible worlds." It is like saying that for all we know, there might be only 7 natural numbers, all of which are prime. This is nonsense: if some things don't at least satisfy

the Peano Axioms, they don't deserve the name "natural numbers." Likewise, there may well be some space of world-like entities with only 7 members; but as these things won't do any of the jobs associated with possible worlds (see Lewis [1986,ch. 1] for a start), they don't deserve the name "possible worlds."

Schwarz's key idea is that if one divorces one's account of modality from all matters conceptual and epistemic, then one's claims are no longer an account of metaphysical modality in the first place. They're an account of something else instead. Putting his point another way, if there is some space of world-like entities that has no epistemological constraints, then we, as philosophers, should have no interest in it. It won't do the jobs we want possible worlds to do. Modal rationalism, with its connections between modality and the a priori, offers one method of providing the desired epistemic constraints.

Kripke (1972) resoundingly defeated the most obvious form of modal rationalism: naive rationalism. But, interestingly, Kripke also laid the groundwork for a more nuanced form of modal rationalism via his discussion of *the factorization view of the necessary a posteriori*, suggested in Kripke (1972). Kripke suggests (in his usual cagey manner) that perhaps all necessary a posteriori truths are derivable from two other truths: one contingent and a posteriori, the other necessary and a priori. For example, consider the necessary a posteriori statement "Heat is molecular motion". Kripke suggests that this statement can be derived in the following way.

(P1) Heat is whatever actually causes heat sensations.

(P2) Molecular motion causes heat sensation.

(C) Therefore: Heat is molecular motion.

(P1) is necessary and (arguably) a priori. Kripke's idea here is that the reference of 'heat' is determined by something like the reference-fixing description "the actual cause of heat sensations". (P2) is contingent and a posteriori. It's contingent because something other than heat might have caused heat sensations (had our brains been wired differently). The conclusion (C) is necessary and a posteriori. According to the factorization view, a derivation of this form can be given for any necessary a posteriori truth.

An a priori implication base has the potential to vindicate the factorization view. The base must implicate all truths, including all necessary a posteriori truths. If (a) the base contains only contingent a posteriori truths and (b) there is an epistemic route from the base to the necessary a posteriori truths that proceeds only through necessary a priori truths, then the factorization view will be vindicated.

Modal rationalism also supports (and gets support from) a connection between conceivability and possibility. However one theorizes about conceivability, it is clearly an epistemic notion. On one account (found in Chalmers (2002) and elsewhere), a proposition P is conceivable if and only if it is not a priori knowable that not-P. According to modal rationalism, the reason conceivability is a guide to possibility is because of the constitutive ties between conceptual matters (and a priori justification in particular) and metaphysical modality. The conceivability-possibility link is a fundamental idea in the vicinity of modal rationalism.

Clearly, then, modal rationalism has its virtues. It prohibits brute or strong necessity, vindicates the factorization view of the necessary a priori, supports a link between conceivability and possibility, and ties together epistemic and meta-

physical space. I think few philosophers would deny at least a weak form of modal rationalism. No one eagerly accepts that penguins exist necessarily and this truth is beyond our (or anyone's) epistemic reach. The important question, then, is how strong the link between metaphysical modality and the a priori is. The master argument requires a particularly strong connection, embodied in the Strong Modal Rationalist Link.

2.5.3 The View From Above

One idea that buttresses the strong modal rationalist picture goes as follows. Imagine an incredibly intelligent and wildly imaginative idealized reasoner positioned outside all the possible worlds. Using her powers of a priori reasoning alone, the reasoner could figure out the entire configuration of modal space. She could describe every possible world down to the finest detail. There is a significant lacuna here. Purely a priori reasoning will not reveal to the thinker whether water is H₂O or XYZ, or whether the concept WATER refers to H₂O or XYZ. So, in a certain sense, she cannot describe every detail of every possible world, or even of our actual world. She can't tell us that our very own Indian Ocean contains water.

This thinker clearly knows a great deal about our world (far more than we will ever know). Using some vocabulary of hers, she can tell us exactly what, at a microphysical and chemical level, the Indian Ocean, or the oceans of Alpha Centauri, contain.²⁶ But she does not know how the 'water' vocabulary applies. In order to know how the 'water' vocabulary applies, she needs to know more about how her token 'water'-representations are related to her environment - are

²⁶The vocabulary in which the reasoner will be able to describe the world will be semantically neutral, because such a vocabulary does not pose the obstacles that 'water' does.

they causally or otherwise linked to H₂O, to XYZ, or to something else?

The idealized thinker can, using her own canonical vocabulary, describe every detail of every possible world.²⁷ In this sense, all of modality is available to a priori reflection. However, she does not know how to describe these worlds using every possible form of representation. In order to know the reference of representations (e.g. one's 'water' representations), one first needs contingent, a posteriori information about how one is situated in relation to one's environment. This is the information the idealized thinker lacks. It explains why she can't tell whether the Indian Ocean contains water. Let's call this general idea, that a priori reflection offers access to all the metaphysical possibilities, but does not always tell one the correct terminology in which to describes those possibilities, *the view from above*. "The view from above" is not just a vista from which one might look out upon modality. It is a claim about the scope of a priori reflection. It captures another element of the modal rationalist picture.

The view from above deflects the significance of the Kripkean necessary a posteriori (and the contingent a priori). Kripke is sometimes thought to have severed the Kantian rationalist connection between metaphysical modality and the a priori. But according to strong modal rationalists, the import of *Naming and Necessity* is more semantic than metaphysical. The entire space of metaphysical possibility is available to a priori reflection. The idealized a priori reasoner can describe all those possibilities using a specialized canonical vocabulary. But she cannot describe those possibilities using any vocabulary whatsoever. This is so because some terms (those that are not semantically neutral) depend for their application on how the actual world turns out. Specifically, they depend on actual

²⁷If there are inexpressible states of the world, then the idealized thinker might not be able to describe those states. But she might be aware of, or know, those states nonetheless.

facts that are knowable only a posteriori.

2.5.4 The Strong Modal Rationalist Link and the View From Above

One can supplement the view from above with additional claims about scrutability, including an ambitious form of the factorization view of the necessary a posteriori. One might claim that the idealized reasoner, given knowledge of the strictly a posteriori facts about the actual world and where she is situated in it, will be able to use her powers of a priori reasoning to learn the truths about that world that were previously beyond her reach: truths involving concepts like *WATER*, *HESPERUS*, and *CICERO*. One might also claim that the idealized reasoner can perform this task from a minimal base of truths - a (neutral) necessitation base, or just the fundamental truths, will do. The strong modal rationalist starts with the view from above and adds on these additional claims about scrutability. \mathcal{P} provides the minimal base of a posteriori facts about the fundamental structure of the actual world (assuming physicalism is true). \mathcal{J} tells the idealized reasoner where, when, and who she is in that world. From there, the strong modal rationalist claims she should be able to figure out the rest.

2.6 Against Strong Modal Rationalism

2.6.1 The Recipe View

Strong modal rationalism posits a tight connection between a priori knowledge and metaphysical modality. This connection entails the truth of the Strong Modal Rationalist Link, which says that any canonically expressed semantically neutral necessitation base, appropriately supplemented with an indexical truth, is also an a priori implication base. Why does the strong modal rationalist answer the supplementation question in this way, i.e. positing the strong rationalist link? In

this section, I'll argue that the strong modal rationalist's answer to the supplementation question is based on a certain meta-semantic view: the recipe view. By "a *meta-semantic* view", I mean a view about how representations come to acquire their meanings.²⁸

According to the recipe view, each term or concept is associated with an a priori "recipe" that determines the reference of the term. For example, the a priori recipe associated with 'water' is something like "the clear wet liquid that flows in the lakes, rivers, and streams and falls from the sky as rain". If H₂O is this clear wet liquid, then water is all and only the H₂O.

The recipe view is not committed to the claim that the a priori referential recipes are captureable via pithy descriptions such as 'the clear wet liquid that flows in the lakes, rivers, and streams and falls from the sky as rain' or 'the teacher of Alexander'. We can model a term's a priori recipe as an infinite set of conditionals. The antecedent of each conditional expresses some possibility for how the world might turn out to be. The consequent states what 'water' refers to should that antecedent be the case. For example, we might model the recipe of 'water' as follows:

- If H₂O is the clear wet liquid that falls from the sky as rain and fills the lakes, rivers, and streams, then water is H₂O.

- If XYZ is the clear wet liquid that falls from the sky as rain and fills the lakes, rivers, and streams, then water is XYZ.

²⁸I use 'meaning' in a generous sense, to cover representational qualities both at the level of reference and at other levels (such as the level of "sense" (cf. Frege (1952/1892a))). I also use 'meaning' to include the representational qualities of any variety of contentful entity, not just linguistic items. If there are non-linguistic representations (which I believe there are), such as perceptual states, they have meaning as well, as I use the term.

- If H₂O is the clear wet liquid that falls from the sky as rain, emerges from faucets, fills the oceans, lakes, and aquifers, and ..., but XYZ and not a drop of H₂O flows in the streams, then water is all and only the H₂O. (In such a case, we'd mistakenly believe, until corrected by chemists, that streams contain water.)
- et cetera...

Because it does not employ descriptions, the recipe view is not obviously susceptible to the anti-descriptivist arguments of Kripke (1972). However, the recipe view can be seen as a brand of "neo-descriptivism". The recipe view is more sophisticated than descriptivism but preserves its spirit. The two-dimensional semantic theories of Chalmers (1996a, 2006a) and Jackson (1998) offer paradigms of the recipe view. Unsurprisingly, Chalmers and Jackson are both proponents of Strong Modal Rationalism as well.

According to descriptivism about proper names, each name is associated with a description that determines the extension of the name. Of course, the description can't determine extension all by itself. Facts about what actually satisfies the description also play a role. But there is clearly a sense in which the description determines reference. The description does the meta-semantic work.

In a similar vein, the recipe view posits a rich variety of a priori connections associated with any term or concept. These a priori connections don't determine reference all by themselves. Instead, they operate as a function that, given a centered world input ("considered as actual"), yield an extension as output. But there is a clear sense in which the a priori recipe determines reference. Henceforth, I'll speak of these recipes as "determining reference", leaving implicit the role of the centered world considered as actual. These a priori connections pro-

vide a type of conceptual role for the term. This conceptual role is in many ways similar to Frege's notion of "sense" (cf. Frege (1952/1892a)). From here, I'll refer to this conceptual role, understood as the concept or term's a priori connections, which can be modeled as an infinite set of conditionals, as the concept or term's sense.

2.6.2 The Trench View

In this section I offer a different meta-semantic picture, the *trench view*, according to which sense, taken to be constituted by a priori conceptual role, does not determine reference. In the next section (2.6.3: "The Recipe View and the Strong Modal Rationalist Link"), I'll use this picture to cast doubt on the strong modal rationalist's answer to the supplementation question. In section 2.7.5 ("Toward a More Moderate Rationalism"), I'll use this alternative meta-semantic model to offer an alternative answer to the supplementation question.

On the alternative meta-semantic picture, *the trench view*, the determination of meaning is a much messier affair than the recipe view supposes. According to the trench view, there's no neat and clean purely a priori reference-determining function. Instead, our terms and concepts are inextricably bound up in the world. They get their meaning in large part through a practice of using those terms, and particularly by demonstratively applying those terms to items in the world. A *demonstrative application* of a term (or concept) occurs when a representational agent demonstrates some object in the world and applies the term (or concept) to that object (e.g. "That is water"). The demonstration and/or application can occur in explicit language (written or spoken), in thought, or in perception.

One can highlight the difference between the recipe view and the trench view using the procedure of "considering worlds as actual". When we normally con-

sider what a word refers to at some imagined possible world, we consider that world *as counterfactual*. For example, we might ask what 'water' refers to at the possible world containing Putnam (1975)'s twin-earth, in which XYZ, not H₂O, fills the lakes, rivers, and streams. When we consider twin-earth as counterfactual, we declare that 'water' picks out all and only the H₂O at twin-earth. But we can also consider what 'water' would have referred to, had twin-earth been actual. Here, we consider twin-earth *as counteractual*. If the actual world had turned out to be like twin-earth, then 'water' would have referred to all and only the XYZ.

We've already seen that the a priori recipes posited by the recipe view can be modeled by an infinite set of conditionals. The antecedent of each conditional corresponds to some way the world might turn out to be. In other words, the antecedent involves a world considered as actual. If the actual world turns out to be thus-and-so, then 'water' refers to H₂O. If the actual world turns out to be this-and-that, then 'water' refers to XYZ.

We can highlight a key difference between the recipe view and the trench view by considering counterfactuals containing no representational agents. According to the recipe view, a term such as 'water' or a concept such as WATER can have a determinate extension at a world considered as counterfactual even if no one has ever used the term 'water' or thought with the concept WATER. Once we have a centered world considered as actual, we simply look for the appropriate line of the a priori recipe. For example, to determine what 'water' refers to at a centered world considered counterfactually, we look in that world for the clear wet liquid that flows in the lakes, rivers, and streams near the center. That liquid is the

water. The world might contain no language users at all.²⁹

In contrast, the trench view denies that, when we consider worlds as counteractual, 'water' (or WATER) has any reference at all if 'water' has never been used by an agent. The trench view maintains that the reference of 'water' and of WATER is highly driven by particular concrete uses of language and other representational states (such as thoughts). If no representation occurs at a world, there is no use of representation to give 'water' a referent. On this picture, the reference of 'water' is not determined by some complicated recipe a priori associated with 'water'. Instead, the fact that language users have labeled H₂O and not XYZ as 'water' makes 'water' refer to H₂O rather than XYZ. Of course, how an agent demonstratively applies their terms is partly determined by the a priori associations that the term has. But it's also determined by the agent's (potentially false) beliefs about the object of demonstration and about the term. Often an agent's demonstrative application is not guided by a strictly a priori recipe. But the demonstration plays a role in securing reference nonetheless.

Demonstrative applications provide a sort of meta-semantic "glue" that functions to tie the reference of a term to the demonstratum. When an agent labels some object as a 'cat', that very act functions to make it more likely that the 'cat' representation type refers to objects of the demonstrated category. This is so even if the labeled entity is not, given the current extension of 'cat', a cat. Each act of demonstrative application has some meta-semantic force, however small, that pushes the reference of the term toward the demonstratum. One mislabeling will almost never alter the extension of 'cat'. In certain circumstances, however, mas-

²⁹Chalmers (2006a): 107-108 is quite clear that primary intensions (which are a form referential "recipe") remain defined, and have determinate extensions, at counterfactually considered worlds with no language users.

sive mislabeling could shift the reference of the term.³⁰

The trench view needn't deny that something like an a priori recipe, conceptual role, sense, or even description plays a role in determining reference. But I'm highly skeptical that each term is associated a priori with a sense or recipe rich enough to determine reference. Furthermore, the meta-semantic glue that demonstrative applications provide can't be reduced to, or replaced by, a priori recipes, dispositions to apply, or beliefs.

I'd like to offer two "just so stories" about reference determination. The stories represent the two meta-semantic approaches already described. The stories admittedly caricature their targets in certain respects. But I think the stories illustrate quite effectively the differences between the two meta-semantic views.

The recipe view offers something like the following picture of reference determination. We start in a dark room, with no contingent knowledge of how the world is. We come up with some term: 'water'. We then think about all the possible ways the world beyond the dark room might be, and decide what 'water' will refer to in each case. If the world is thus and so, water is H₂O. If the world is that and such, water is XYZ. Et cetera. Each line of the recipe corresponds to a different way the world could turn out to be. Then we leave the dark room and check how the world actually is. It turns out that H₂O is the wet clear liquid in the lakes, rivers, and streams. Voila! Water is H₂O. (In fairness to proponents of the recipe view, the recipes need not be completely decided in advance. The

³⁰Evans (1973) discusses two cases that seem to demonstrate the phenomenon of continued demonstrative application functioning to meta-semantically glue a term (a name) on to an object to which the term did not originally refer. Evans (195-6) notes that the term 'Madagascar' originally referred to a portion of mainland Africa. But Marco Polo initiated a process of misapplication - demonstrative application - of the term to the large island we now know as 'Madagascar', eventually causing the term to genuinely refer to that large island. Evans' 'Turnip' case (206-7) involves a similar shift in reference through demonstrative (mis)application.

recipes can be implicit in the dispositions of language users. The main claim of the recipe view is that the reference-determining work is done cognitively, via the multi-line recipe. The job of the world is limited to determining which line of the recipe to use.)

The trench view offers a much messier picture of reference determination. We don't start in a dark room outside the world and decide in advance what our terms refer to, even in conditional form. Instead we start in the world, with all its pressures and demands bearing down on us. A *Tyrannosaurus Rex* crashes through the bushes, snorting and gnashing its teeth. We point, yell, 'Danger!' and run. After making our escape, we discover a stream of clear cool water to slake our thirst. This substance is useful and important. We'll need to identify it again and communicate about it. We introduce a label: 'water'. "This is water," we declare. Of course, we have various beliefs about water that guide our application of the term. Perhaps we believe water to be the clear wet liquid that flows in streams. But the reference of the term is driven both by these beliefs and our dispositions to apply the term and by the history of using the term as a label for certain stuff. Importantly, our continued use of the term 'water' as a label plays a role in reference-determination that can't be reduced to any recipe, beliefs, or dispositions. And it can turn out that water is neither clear, nor wet, nor flows in the streams.

On the trench view, demonstrative applications play a vital role in determining reference. Of course, we need other representational or intentional abilities to lock our demonstrations on to objects in the world. But our ability to "lock on" to some particular in the world is a somewhat primitive ability, usually facilitated by the perceptual system. In order to demonstrate something, we don't need

a description satisfied uniquely by the demonstratum. The simpler lower-level representational capacity to lock on or demonstrate items in our perceptual field facilitates the generation of the higher-level representational capacities indicative of linguistic and conceptual representation. The ability to represent something as water using the term 'water' or the concept WATER is such a higher-level capacity.

This important role of demonstrative applications is discussed in Burge (1986): 716. He speaks of "applications" of terms: "ordinary, indexically aided, perception-based uses". 'That is water / a cat / Madagascar' are paradigm examples. Burge states that "ordinary words (or notions) for ordinary entities are given their life and meaning through such applications." I whole-heartedly agree. The recipe view's main mistake is to ignore the important role of these demonstrative applications, and the degree to which the generation of meaning is a living and dynamic affair, tied to representational activity and action on the ground, not decided cognitively in advance via pre-planned recipes (however implicit in the dispositions of agents these recipes may be).

Putnam (1962)'s robotic cat thought experiment seems to tell against the recipe view and in favor of the trench view. Putnam imagined a scenario in which the creatures we'd been calling 'cats', invited to warm our laps and made members of our home, were robots all along. Putnam persuasively argued that if this turned out to be so 'cat' would refer to a certain type of robot, and not to any animals. If there is an a priori recipe associated with 'cat', it likely includes features like "furry", "animal", and "purrs". But Putnam showed that, through the demonstrative application of 'cat' to a certain type of robot, 'cat' can come to refer to something that is neither an animal, nor furry, nor purrs. In this case at least, the work of reference determination appears to be done by demonstrative appli-

cations, not recipes or a priori associated features. Demonstrative applications can even overcome and outweigh a recipe that pulls reference in the opposite direction (e.g. toward furry animals). The trench view takes the phenomenon demonstrated by Putnam's cat scenario to exemplify the generation of meaning in a large class of cases. Demonstrative applications do much of the work in securing both meaning and reference.

The recipe view overestimates the degree to which the generation of meaning is an intellectual or cognitive affair. The recipe view claims that users' intellects determine a conditional recipe, decided away from the world, that determines meaning. The role of the world is limited to telling us which conditional, which line of the recipe, to use. The alternative trench view holds that the generation of meaning occurs down in the trenches (hence "the trench view"). Agents use representations in their attempts to survive and thrive in the muck and helter-skelter of the world. Their assignment of meaning to representational tokens is done in a haphazard manner that fits their integration into and interaction with the world. In sum, the generation of meaning is messier and more world-bound than the recipe view or the strong rationalist suppose.

2.6.3 The Recipe View and the Strong Modal Rationalist Link

The recipe view fits nicely with the strong modal rationalist's answer to the supplementation question. First, the recipes are centered. Water is not just the clear wet liquid that flows in the lakes, rivers, and streams. It's the clear wet liquid that flows around here. The fact that XYZ flows in the lakes, rivers, and streams, on the far side of the galaxy matters not to the extension of 'water'. The fact that the recipes are centered fits conveniently with the strong modal rationalist's demand for an indexical element in the a priori implication base. The indexical element

indicates where “here” is, such that one should investigate the lakes, rivers, and streams nearby.

The recipe view lends support to modal rationalism in another way. The recipes aren't just recipes for extension. They're also recipes for intension. The 'water'-recipe says that if H₂O is the clear wet liquid in the lakes, rivers, and streams around here, then the extension of 'water' includes all the actual H₂O. But the recipe also says if H₂O is the clear wet liquid... , then 'water' picks out H₂O at every possible world. In other words, the recipe says that if the actual clear wet liquid... is H₂O, then water is necessarily H₂O. 'Water' has an intension, a function from worlds to extensions, that picks out the H₂O at every world. The actual world extension of 'water' is a special case.

The intension of 'water' tracks the reference of 'water' across all possible worlds, and thus determines what is and is not metaphysically possible for water. If a reasoner knows the intension of 'water', she knows what is and is not metaphysically possible for water. On the recipe view, one can know a priori the intension of 'water' conditional on how the actual world is. The information in the necessitation base tells one how the actual world is. From this information, one can determine a priori what the intension is, allowing the reasoner a priori access to all of modal space - the rationalist dream.

Let's now explore how a priori recipes facilitate the strong modal rationalist answer to the supplementation question, which we see in the strong modal rationalist link.

- **(Strong Modal Rationalist Link):** A canonically expressed semantically neutral necessitation base, supplemented with an indexical truth, a priori implicates every proposition it necessitates.

Imagine an ideal and powerful reasoner, who has been given an indexicalized semantically neutral necessitation base \mathcal{F} . \mathcal{F} describes, in canonical neutral vocabulary, the complete layout of the world's fundamentalia: all the fundamental particles (e.g. quarks, leptons, and bosons), their fundamental properties (e.g. mass, spin, charge), and some fundamental laws that govern the development of these fundamentalia over time. These fundamentalia might be physical, or they might not. For example, they might include consciousness particles or a fundamental psycho-physical law. Let \mathcal{F} be supplemented with an indexical truth \mathcal{J} that locates the ideal reasoner in space-time. According to the strong modal rationalist link, $\mathcal{F}\mathcal{J}$ is an a priori implication base - it a priori implicates every truth. Roughly, this entails that the ideal reasoner, informed that $\mathcal{F}\mathcal{J}$ is the case, can reason a priori to knowledge of all truths.³¹ Can she do it?

There is some reason to be optimistic. After all, \mathcal{F} describes the world's fundamental layer. This layer is metaphysically responsible for the rest of the world's contents. It also necessitates the complete state of the world. The facts about water, koalas, and Africa are necessitated and depend upon the truths about fundamentalia, the very truths to which the ideal reasoner has access. In some sense, all the information is already there.

$\mathcal{F}\mathcal{J}$ does not contain any concepts like WATER, KOALA, or AFRICA. \mathcal{F} is expressed in fundamental vocabulary, using concepts like ELECTRON, LAW, FORCE,

³¹Following Fitch (1963), there are good reasons for thinking that in fact it is impossible to know all truths. Let P be any truth that is not currently known. If P is not known, then " P and no one knows that P " is true. Call this truth F . F seems to be unknowable. If it were known, it would become false, and thereby not known (falsehoods can't be known). However, the issue can be side-stepped. What is really at issue is whether $\mathcal{F}\mathcal{J}$ a priori implicates every truth, i.e. whether, for every truth T , a material conditional of the form $\mathcal{F}\mathcal{J} \rightarrow T$ is knowable a priori. The conditional can be knowable a priori even when T is not. I'll usually leave issues about unknowable truths suppressed and instead speak in the easier, if somewhat sloppier, mode, of "coming to know" all truths. For further discussion of the Fitchian obstacles to a priori implication cf. Chalmers (2012): Chapters 2 and 4.

mathematical concepts, and perhaps some mental concepts (if fundamentality dualism is true). But the a priori reasoner must be able to reason a priori from \mathcal{F} to truths containing the concepts WATER, KOALA, and AFRICA. Consider the proposition THE LARGEST BODY OF WATER IN AFRICA HAS A SURFACE AREA OF MORE THAN 50,000 SQUARE KILOMETERS. This proposition is true. (Lake Victoria has a surface area of 68,800 km.²).³² But how might the reasoner come to know a priori from \mathcal{F} ? First, she needs to determine what WATER and AFRICA refer to. The a priori referential recipes associated with these concepts will help her.

For ease of exposition, we'll assume that physicalism is true. \mathcal{F} is expressed in the vocabulary of fundamental physics, math, and logic. The idealized reasoner starts from the centered description \mathcal{F} of the world at its most fundamental level. The a priori implication of all truths starts from the very small fundamentalia and proceeds up, piece-meal, to the macroscopic. The reasoner starts by using the recipes a priori associated with the concepts of chemistry, which are either written in the vocabulary of micro-physics or in some topic-neutral vocabulary, to deduce a priori all the facts about chemistry. She then uses the information about microphysics plus chemistry, along with the recipes of micro-biological concepts, to deduce the micro-biological facts. Repeat this process, moving up the chain of a priori implication from the micro to the macro.

The reasoner gradually fills in her picture of the world's contents, proceeding from small to big. Eventually, the idealized reasoner will discover that H₂O, not XYZ, is the clear wet liquid that flows in the lakes, rivers, and streams around here and which falls from the sky as rain. The reasoner will conclude that H₂O, and not XYZ or some other substance, is water. She'll similarly use the recipes

³²Cf. Wikipedia (2013).

associated with AFRICA and other concepts (or terms), to determine the reference of those concepts (or terms). Once she knows the reference of all these concepts, she can evaluate which propositions are true and which false. She'll learn that the largest body of water in Africa does have a surface area of more than 50,000 square kilometers.

One learns that 'The largest body of water in Africa has a surface area of more than 50,000 square kilometers' is true by learning that the reference of 'The largest body of water in Africa' is the largest body of H₂O in the continent containing the Sahara and the Congo, which one learns by determining what 'water' and 'Africa' refer to.

The basic connection between the strong modal rationalist link and the recipe goes as follows. According to the recipe view, the reference (and intension) of a term or concept can be read off from the a priori recipe in combination with a centered world. The centered world tells one which line of a priori recipe to use. The strong rationalist link says that a semantically neutral indexically supplemented necessitation base a priori implicates all truths. The semantically neutral indexically supplemented necessitation base provides a centered world.³³ The reliance on a priori implication gives one recourse to the a priori recipe. From these ingredients, one can determine the reference of all terms, and come to know all truths.

Contrast the picture of a priori implication that flows from the recipe view

³³The route from the recipe view to the strong rationalist link is not as simple as I have suggested. There is no guarantee that the a priori recipes are written in the same (likely fundamental) vocabulary as the necessitation base \mathcal{F} . In other words, the centered world description \mathcal{F} might not work as input into the recipe view's a priori reference-determination function. Thus, even if the recipe view is true, the strong rationalist link might still fail. However, I wish to ignore this further problem for the strong rationalist link. Instead I focus on denying the recipe view itself, rather than the connection between it and strong modal rationalism.

with the picture that flows from the trench view. According to the trench view, terms might have a priori recipes, but they do not, in combination with a centered world, determine reference. A priori associations could play some role in restricting the reference. For example, one might know a priori that water isn't a type of sandwich. But the a priori associations leave space for other reference-determining factors.

According to the trench view, the slack in the project of reference-determination is picked up by the process of language (and concept) users interacting (often causally) with items in their environment. In particular, language users demonstratively apply their terms (and concepts) to things they encounter. Thus, a semantically neutral description of the world, even supplemented with an indexical truth, will often leave one unable to determine what terms such as 'water', 'cat', and 'red' refer to. Before one can do so, one needs also to know what types of things users have demonstratively applied 'water', 'cat', and 'red' to.

2.7 The Demonstrative Reply to the Master Argument Against Physicalism

In this section, I describe the *the demonstrative reply* to the Master Argument Against Physicalism. According to the demonstrative reply, the failure of a priori implication from the physical truths to the truths about consciousness stems from a crucial and ineliminable meta-semantic link between phenomenal concepts and terms (such as RED_{ph} /'red_{ph}' and $PAIN_{ph}$ /'pain_{ph}')³⁴ and demonstrative appli-

³⁴A *phenomenal concept* applies to experiences and characterizes an experience according to what it is like to be the subject of that experience. For example, RED_{ph} (the subscript 'ph' denotes 'phenomenal') is the phenomenal, experiential, concept of red. RED_{ph} applies to experiences with a certain qualitative character (you know the one). Neither light-waves nor objects can be red_{ph}. Tomatoes, fire trucks, and strawberries are red, but they cause red_{ph} sensations. Analogous comments apply to *phenomenal terms*, such as 'red_{ph}' and 'pain_{ph}'. The English term 'red' is likely

cation of those concepts and terms to conscious experiences. The failure of a prior implication is an expected consequence of this meta-semantic link and is not symptomatic of the presence of mentality in the fundamental bedrock of reality.

2.7.1 The Demonstrative Reply

The crucial move in the Master Argument Against Physicalism is the inference from an epistemic a priori gap between the physical and consciousness to a metaphysical gap that entails dualism. The supplementation question asks, "What must be done to a necessitation base to create an a priori implication base?" An answer to the supplementation question creates the framework for squeezing metaphysical results about necessitation and fundamentality from epistemic claims about a priori implication.

The Master Argument relies on the strong rationalist answer to the supplementation question, and its robust connection between modality and the a priori, embodied in the strong modal rationalist link. According to the strong rationalist link, a necessitation base expressed in semantically neutral vocabulary need only be supplemented with an indexical truth to become an a priori implication base. Because the physical truths, expressed in a neutral fashion and supplemented with an indexical truth, do not a priori implicate the truths about consciousness, the physical truths do not form a necessitation base. Thus they do not express the fundamental state of the world. Fundamentality physicalism is false. Or so the master argument goes.

We might ask why a necessitation base needs any help at all a priori implicat-

polysemous between, or a mongrel of, the red-terms that apply to light waves, to surfaces, and to sensations.

ing all truths. The basic reason is that necessitation (and fundamentality) is, first and foremost, a property of states of affairs, not a property of representations. Certain representations (BOB IS A BACHELOR) necessitate other representations (BOB IS NOT MARRIED) because of relations between the objects, properties, and states of affairs those representations denote. A priori implication, however, is solely a relation between representations. The master argument seeks to use a priori implication as a guide to necessitation and fundamentality. We must take care when we shift the locus of inquiry from the things themselves to our representations of them. The move to the representational level introduces complications. We've seen several of these complications: the need to express truths in a canonical vocabulary and the need to include an indexical truth.

I argued earlier (section 2.3.4: "Naive Modal Rationalism") that the indexical truth must be included in any a priori implication base in order to implicate indexical truths (such as I LIVE IN LOS ANGELES) and even certain third-personal truths (such as WATER IS H₂O). These are reasons associated with success. In order to *succeed* as an a priori implication base, any such base must include an indexical truth. But let us inquire more deeply into the need for the indexical truth. Why does an a priori implication base need indexicals to succeed?

The reason a priori implication bases must include indexical truths is that indexicality, our perspective on the world, is a representationally basic feature. A priori implication bases must include representationally basic features. But the inclusion of a representationally basic feature in an a priori implication base does not entail that this feature is metaphysically basic (i.e. metaphysically fundamental). The need to include I AM IN LOS ANGELES in the a priori implication base does not indicate that I, or Los Angeles, lie in the bedrock of reality.

Before we can draw conclusions about fundamentality from claims about a priori implication, we must first carefully investigate which features of a priori implication bases are *merely* representationally basic. The indexical truth is one such feature. *The demonstrative reply* to the master argument against physicalism claims that demonstratives are similarly representationally basic. As a result, any a priori implication base must include some demonstratives. Furthermore, the demonstrative reply claims a vital connection between our representations for conscious experience (phenomenal concepts and terms) and demonstratives. This connection prohibits the physical truths (even when semantically neutral and supplemented with an indexical truth), when not aided by demonstrative truths, from a priori implicating the truths about consciousness.

According to the demonstrative reply, the connection between a priori implication and necessitation is not as tight as the strong modal rationalist claims. The strong modal rationalist, relying on the recipe view of how our representations acquire their meanings, ignores the crucial meta-semantic role of demonstratives and underestimates how much assistance a necessitation base requires to become an a priori implication base. Because of the vital role that demonstrative applications play in the determination of reference, the recipe view is false. As a result the strong modal rationalist link, which underlies premise (P2) of the master argument, fails. Any a priori implication base must include some demonstrative truths. The demonstrative reply denies premise (P2) of the master argument (this premise says that if the physical truths, indexically supplemented, fail to a priori implicate all truths, they fail to necessitate all truths).

Many views can deny the strong modal rationalist link. But the demonstrative reply offers a diagnosis of where the link goes wrong. The link fails to acknowl-

edge the vital meta-semantic role that demonstratives play. Without demonstrative truths in the a priori implication base, we can't determine the reference of our terms (concepts) or the truth values of our sentences (propositions). This is why a priori implication from a demonstrative-less base fails.

Furthermore, the demonstrative reply provides a physicalistically respectable explanation of the epistemic gap between the physical and consciousness. The epistemic gap arises from (i) the representationally basic nature of demonstratives and (ii) the essential and ineliminable connection between demonstratives and our representations of consciousness. (More on this "essential and ineliminable connection" in section 2.7.3: "The Ineliminable Connection Between Phenomenal Concepts and Demonstratives"). The epistemic gap does not stem from the fact that the fundamental features of our world include consciousness itself (a form of dualism).

The demonstrative reply offers the tools to agree with much of the thinking behind the Master Argument while resisting the final step to dualism. Proponents of the demonstrative reply can admit that the strong rationalists are correct about the failure of a priori implication between the physical and consciousness. Strong rationalists claim that consciousness truths (or something of their ilk) must be added to any a priori implication base. Proponents of the demonstrative reply can agree with this claim as well. Consciousness must be included because (a) demonstratives must be included in any a priori implication base and (b) there is a tight connection between demonstratives and consciousness.

Strong rationalists infer from the need to include consciousness in the a priori implication base to a need to include consciousness on the roster of fundamentalia. The demonstrative reply resists this final step. The need to include con-

consciousness stems from the need to include demonstratives, which are representationally basic, and not from the metaphysically basic (i.e. fundamental) nature of consciousness itself.

Lastly, the demonstrative reply, unlike other anti-rationalist replies that deny (P2), is amenable to modal rationalism (though not to Strong Modal Rationalism). In fact, the demonstrative reply naturally points the way toward a weaker form of modal rationalism, to be discussed in section 2.7.5 (“Toward A More Moderate Rationalism: the Moderate Rationalist Link”).

2.7.2 Comparison to other Demonstrative Oriented Replies

The demonstrative reply is a version of *the phenomenal concept strategy*. The phenomenal concept strategy locates the explanatory gap between the physical and consciousness in special features of our concepts for conscious experience (phenomenal concepts) rather than in a metaphysical gap resulting from dualism. There are a large variety of phenomenal concept strategies. I will not canvas them all here.

Many phenomenal concept strategists have noted the special or direct way in which our phenomenal concepts pick up on experience itself (Balog (2012), Loar (1990), Papineau (1998)). Chalmers (2003b), while not a proponent of the phenomenal concept strategy, writes of an “experience... that is being attended to or taken up into the concept” (239). Others have suggested that phenomenal concepts are, in some way or other, demonstrative (Loar (1990), Hawthorne (2002)). My demonstrative reply follows in this track, but there are important differences between my demonstrative reply and existing demonstrative accounts of phenomenal concepts.

The most important difference between my demonstrative reply and exist-

ing demonstrative accounts is that, on my account, the role of demonstratives is strictly meta-semantic. The difference can be expressed using the Fregean notion of *sense* (Frege (1952/1892a)). Other accounts (e.g. Loar (1990): 87) suggest that demonstratives enter into the sense of a phenomenal concept. In contrast, on my view demonstratives play a vital role in making it the case that phenomenal concepts have the sense and reference that they do, but demonstratives need not enter into the sense itself. My story about phenomenal concepts is *meta-semantically demonstrative*. In contrast, many other accounts are *semantically demonstrative*.

I'll attempt to explain the difference between semantically and meta-semantically demonstrative concepts. Suppose I introduce a term by stipulating that 'whapow' is to mean *this sound* (slapping my hands against my thighs). I thereby generate a semantically demonstrative term 'whapow'. 'Whapow' functions similarly to Kaplan (1970)'s 'dthat', which creates a term out of a demonstration. 'Dthat[...]' (directly) denotes whatever is demonstrated at the time of utterance. A description of the demonstration appears inside the brackets [...]. My stipulation of the meaning of 'whapow' is much like a use of 'dhat'. I could have introduced the term by saying, "Whapow = Dthat[slapping my hands against my thighs and gesturing toward the noise]". Most terms are not semantically demonstrative. 'Water', 'cat', 'koala', 'friend', and 'Africa' do not function like 'whapow' or 'dthat[...]'.

But, according to the trench view, many non-semantically demonstrative terms are meta-semantically demonstrative, including 'water', 'cat', and, most importantly, 'red_{ph}'. They are meta-semantically demonstrative because demonstrative application plays a vital role in the determination of reference for those terms.³⁵

³⁵Putnam (1975) discusses the phenomenon of using meta-linguistic/meta-semantic information to introduce kind terms. Putnam, however, fails to distinguish between terms which are semantically vs. meta-semantically demonstrative, indexical, or meta-linguistic. Burge (1982) makes points similar to mine about kind terms in general (not just phenomenal terms) being

Semantically demonstrative accounts of phenomenal concepts run into a familiar problem: it's just implausible that phenomenal concepts (terms) are demonstrative concepts (terms) (cf. Chalmers (2003b) for detailed criticism). 'Red_{ph}' seems quite unlike 'whapow'. As an example, consider Hawthorne (2002)'s demonstrative account of Mary's epistemic progress upon leaving the black and white room and experiencing phenomenal red for the first time. According to Hawthorne, Mary knows before leaving the black and white room that a ripe strawberry will give her a red experience. When she actually sees the strawberry, she comes to know that a ripe strawberry will give her *this* experience. Mary's epistemic progress is constituted by her gaining of demonstrative knowledge of something she already knew.

However, this account seems to grossly underestimate Mary's epistemic progress. There are no rhinoceroses in the black and white room, but we can suppose that Mary knows all about them. Further, we can suppose that Mary has never seen a rhinoceros or been in a position to demonstrate one. When she is released from the black and white room into the safari enclosure at the San Diego Zoo, she sees her first rhinoceros and comes to know that *this* is a rhinoceros. This type of epistemic progress strikes most as unimpressive and vastly different from Mary's epistemic progress with respect to phenomenal red.³⁶ But according to the

meta-semantically demonstrative in the sense explained.

³⁶It is difficult to make out exactly what the difference between Mary's epistemic progress in the two cases is. In both cases, Mary can't, from the black and white room, have the relevant demonstrative knowledge ("this is red_{ph}" or "this is a rhinoceros"). But Mary, and proponents of the knowledge argument, should know enough about demonstratives to expect this type of epistemic gap. If that were all there was to what Mary didn't know, the knowledge argument would not have been as influential as it has been.

One way to make out the difference is by imagining Mary with a "phenomenal color palette" containing the various shades of phenomenal color experience. Mary is tasked with painting the world with this palette, painting each object with the type of color experience it tends to cause. It's plausible that she won't know whether to paint the strawberry red_{ph}, blue_{ph}, or yellow_{ph}. (Let's

demonstrative account, they're on a par. This strikes many, including myself, as implausible.

Semantically demonstrative accounts have a nice story about the epistemic gap between the physical and consciousness. In the philosophy of a priori implication, it is well understood that exceptions must be made for demonstratives (cf. Chalmers (2012): 285-7). In the same way that one can't get first-person indexical truths from third-person truths, one can't get demonstrative truths either. Demonstrative truths are representationally basic in the manner described in the previous section. If the truths about consciousness employ phenomenal concepts, and phenomenal concepts are demonstrative, then the truths about consciousness are demonstrative. One should not expect a priori implication of demonstrative truths from non-demonstrative truths (such as the physical truths \mathcal{P}). Chalmers (2012): 285-7 seems to agree with this point (though not with the claim that phenomenal concepts and truths are semantically demonstrative). He places demonstrative truths directly in his a priori implication base. But on Chalmers' picture, the role of demonstrative truths appears to be limited to facilitating the a priori implication of other demonstrative truths.

Semantically demonstrative accounts pay a steep price for their nice story about the epistemic gap. They commit themselves to an implausible account of the content of phenomenal concepts. Phenomenal concepts seem quite unlike the concepts "that chair" or "that[man in a dark coat]". Mary's epistemic progress seems more substantive than the move from "tomatoes will cause a red_{ph} experi-

imagine that Mary has the phenomenal palette containing the colors, but not names for the colors on the palette. She can't use those terms to facilitate her labeling). If this is correct, there is an epistemic gap between the physical truths and the red_{ph} truths that goes beyond the inability to demonstrate red_{ph} while in the black and white room. There is no analog of this more substantive epistemic gap in the rhinoceros case.

ence” to “tomatoes will cause this experience.”

My meta-semantically demonstrative account aims for the best of both worlds. Because of the connection between phenomenal concepts and demonstratives, we should expect an epistemic gap between the phenomenal truths and the physical truths and not draw anti-physicalist conclusions from it. But because the connection to demonstratives is meta-semantic rather than semantic, my demonstrative reply avoids the unsavory and implausible claims about the content, or sense, of phenomenal concepts to which semantically demonstrative accounts are committed.

To reap the benefits of semantically demonstrative accounts (i.e. a nice physicalistically respectable story about the epistemic gap), a meta-semantically demonstrative account requires legwork above and beyond that required for the semantic approach. Semantically demonstrative accounts get the simple explanation of the epistemic gap: “Phenomenal truths are demonstrative. Demonstrative truths are special. They aren’t a priori implicated by physical truths. Failure of a priori implication stems from this fact about demonstratives, not dualism. End of story.” My demonstrative reply requires explaining why the meta-semantic dependence of phenomenal concepts on demonstratives entails a failure of a priori implication. The trench view provides this explanation. Demonstratives play a vital role in the determination of reference. They play a particularly important role with phenomenal concepts (more on this shortly in section 2.7.3). If you don’t know what ‘red_{ph}’ has been applied to in the past, you can’t know what ‘red_{ph}’ refers to, and you can’t figure out the ‘red_{ph}’ truths. Because the demonstrative applications upon which the reference of ‘red_{ph}’ depends do not appear in the physical a priori implication base, that base fails to a priori implicate the

truths about red_{ph} . This story about the epistemic gap between the physical and consciousness falls directly out of the meta-semantically demonstrative account of phenomenal concepts, and in no way supports metaphysical dualism.

2.7.3 The Ineliminable Connection Between Phenomenal Concepts and Demonstratives

In this section, I will explain the previously mentioned sense in which our representations of consciousness, particularly phenomenal concepts, have an “essential and ineliminable connection” to demonstrative applications. The point is best made by considering a rejoinder to the demonstrative reply to the Master Argument. (I’ve encountered this rejoinder many times in my discussions with strong modal rationalists).³⁷

The demonstrative reply relies on the trench view, according to which demonstrative applications of concepts and terms plays a vital role in determining reference. The trench view is a plausible meta-semantic approach to natural language and to our conceptual cognitive capacities. However, imagine the following response on behalf of fans of the Master Argument. “Perhaps the trench view is correct as a meta-semantic theory of natural language and of the concepts limited reasoners such as ourselves employ. But questions about a priori implication are to be decided by considering all the types of reasoners, languages, and concepts there could be. We must look beyond our cognitive limitations. The recipe view is true for some language and conceptual scheme, employed by beings with cognitive abilities far beyond ours. If so, then dependence on demonstrative application is irrelevant, because this dependence is a contingent feature of natural

³⁷Thanks are due to David Chalmers, Edward Elliott, Kelvin McQueen, and Wolfgang Schwarz for helpful discussion.

language and of our conceptual abilities, rather than a necessary feature of representational capacities more generally.”

This line of thought is intriguing. Let’s flesh it out a bit more. Let’s call the envisioned language (and conceptual scheme) utilized by powerful God-like reasoners who don’t have a need for demonstrative applications to secure the reference of their terms *the god language*. The gods who use this language are tremendously imaginative and have vast computational abilities. Their language contains a term ‘water^g’. ‘water^g’ is the god-language version of ‘water’. It refers to all and only the water.

The gods are meticulous in their creation of language. When they create a term, they imagine every possible way the world could be, and state what the term would refer to should each of those possibilities turn out to be actual. The a priori recipes of the god-language aren’t implicit in their dispositions. The gods explicitly lay out the recipes, perhaps inscribing them in a lawbook of meaning. (There are questions about how even gods might carry out this infinite task. Let us ignore such complications.) The dialectic here will operate mostly at the level of language, but we could adopt a similar understanding of god-concepts.

How does the possibility of a god language, for which the recipe view seems the correct meta-semantic theory, increase the plausibility of a priori implication from a demonstrative-less base? Because the god-language seems not to meta-semantically depend on demonstrative application, and its terms have explicitly laid out a priori referential recipes (perhaps even written in fundamental vocabulary), a priori implication of the god-language truths won’t run into the demonstrative-oriented problems we saw in section 2.6.3 (p116). For example, the a priori recipe associated with ‘water^g’ has a line that explicitly says, “If \mathcal{F}

is the case, then the water is all and only the H_2O (and necessarily so)^g.” With this a priori recipe and $\mathcal{F}\mathcal{J}$ in hand, one does not need to first know what one has demonstratively applied ‘water^g’ to in the past in order to learn the reference of ‘water^g’. Thus, it appears that the demonstrative obstacle to a priori implication of the god language truths has been removed. Truths such as ‘water is H_2O ^g’, ‘cats are animals^g’, and ‘The largest body of water in Africa has a surface area of more than 50,000 square kilometers^g’ seem to be a priori implicated by the demonstrative-less base $\mathcal{F}\mathcal{J}$. However, the truths containing natural language terms such as ‘water’, ‘cat’, and ‘Africa’, are not obviously a priori implicated in this fashion.

Let’s suppose that the result of resort to the god-language is a priori implication of a god-language description of the world. The gods will know a great deal about water, cats, and Africa. But their knowledge will not proceed via the concepts WATER, CAT, and AFRICA. Instead it will involve the concepts WATER^g, CAT^g, and AFRICA^g. In that sense, the resulting description is incomplete. It fails to include truths involving WATER. But in another sense the description is complete, or close enough. The gods can describe, in excruciating detail, the state of the world. They may even be able to describe the truths about consciousness, using concepts such as RED_{ph}^g and PAIN_{ph}^g.

I have two responses to the attempt by fans of the master argument to avoid the problems for a priori implication arising from meta-semantic dependence on demonstratives by resort to the god-language (or anything like it). First, I wish to cast doubt on the idea that a god-language description of the world can function as a genuine substitute for a description in ordinary language. There’s a sense in which the god-language description is light-weight, and its a priori implica-

tion from \mathcal{FJ} unimpressive. Second, and more importantly, one cannot avoid the meta-semantic dependence of phenomenal concepts and terms on demonstrative applications to conscious experience. The relationship between the two is ineliminable.

Let us evaluate how the possibility of this god-language affects the dialectic. The a priori implication from \mathcal{FJ} of the WATER^g , CAT^g , and AFRICA^g truths seems hard to deny. However, one should question whether the WATER^g , CAT^g , and AFRICA^g truths are on equal standing with the WATER , CAT , and AFRICA truths. There's reason to be skeptical that a world-description in terms of WATER^g , CAT^g , and AFRICA^g is genuinely complete, or as informative as a description that uses WATER , CAT , and AFRICA .

Consider the following example, revolving around the glass jar currently sitting on my desk. The gods introduce a term, 'glass jar^g', and stipulate all the conditions under which a glass jar^g exists, and the conditions under which 'there is a glass jar on the desk^g' is true. As it turns out, glass jar^gs exist exactly when glass jars exist, and there is a glass jar^g on the desk^g exactly when there is a glass jar on the desk. Consider a god who has managed to deduce a priori from \mathcal{FJ} that there is a glass jar^g on the desk^g. The god does not know the proposition $\text{THERE IS A GLASS JAR ON THE DESK}$, but he knows a nearby proposition: $\text{THERE IS A GLASS JAR ON THE DESK}^G$.

I'm skeptical that knowledge that there is a glass jar^g on the desk^g is "just as good" as knowledge that there is a glass jar on the desk. 'There is a glass jar on the desk' is a macroscopic truth about the world. 'There is a glass jar on the desk^g' seems more like a complicated microscopic truth. Glass jars^g are stipulated to exist whenever certain micro-arrangements of particles occurs. \mathcal{FJ} is such an

arrangement. The transition from knowledge of $\mathcal{F}\mathcal{J}$ to knowledge that there is glass jar on the desk seems substantive, whereas the transition to there is a glass^s jar on the desk^s seems not. If I stipulate 'gleeb' as a term for any glass jar located on the desk, and then infer from the fact that there is a glass jar on the desk in my office to the fact that there's a gleeb in my office, I have not made much (any?) epistemic progress. The gods who infer from $\mathcal{F}\mathcal{J}$ to the fact that there is a glass jar^s on the desk^s engage in a similarly weak form of epistemic progress. In sum, I'm skeptical that a priori implication of a world-description in terms of WATER^s, CAT^s, and GLASS JAR^s is actually a different method of describing the water, cats, and glass jars, rather than a re-hashing of $\mathcal{F}\mathcal{J}$.

While I am doubtful that a priori implication of truths in the god-language yields genuine epistemic progress, my second, and primary, response to the god-language strategy does not rely on this idea. Instead, I claim that it is impossible to have a phenomenal concept (or term) whose reference does not ultimately track back, in some way, to demonstrative application. It might be possible to get the truths about water, cats, and Africa, via the concepts WATER^s, CAT^s, and AFRICA^s, all of which have explicitly laid out a priori referential recipes and don't rely on past demonstrative applications for their reference. But any genuinely phenomenal term (concept) will rely on past demonstrative applications, either of that term (concept) or another phenomenal term or concept.

It's easy enough to create a term that refers to a type of experience without resorting to demonstratives. I hereby introduce the term 'gluub', which I stipulate to refer Robert Zimmerman's favorite type of color experience. But 'gluub' is not a genuinely phenomenal term (like red_{ph} or pain_{ph}), because it isn't associated with conscious experience in the right way. For starters, there's no phenomenality

- no redness or the like - in the term's sense. Eventually, to lock one's phenomenal term (concept) on to its referent, one must point and say, "That's the type of experience I'm talking about."

Imagine the gods attempting to create a god-language term 'red_{ph}^g', without relying at all on demonstratives. They might describe red_{ph}^g as the type of experience between orange_{ph}^g and violet_{ph}^g. It's doubtful that this works, but even if it did, it merely passes the buck. 'Orange_{ph}^g' and 'violet_{ph}^g' will either get their reference secured (at least partly) through demonstrative application ("This is orange_{ph}^g!") or rely on other terms whose reference was secured demonstratively. Eventually, phenomenal terms must meta-semantically track back to, and rely on, demonstrative application to a conscious experience.

If this is correct, then even in the god-language, a priori implication of the phenomenal truths is only possible from a base that records the demonstrative applications upon which phenomenal terms depend. One can't avoid the need for demonstratives by switching to the god-language or something like it.

The resulting picture yields a difference between representations for conscious experience and other representational capacities. Our representations for water, cats, and Africa, meta-semantically rely on demonstratives just as much as our representations for consciousness (i.e. phenomenal terms and concepts) do. But the meta-semantic link between our representations for conscious experience (i.e. phenomenal terms and concepts) and demonstratives is *ineliminable*. We, or some god-like creatures, might be able to get on to the cats, water, and Africa, without using demonstratives. But the realm of conscious experience can't be accessed in the same way. Consciousness must be demonstrated.

2.7.4 Phenomenal Concepts, Demonstration, and the Experiential Requirement

These comments on the ineliminable connection between phenomenal concepts and demonstrative applications to conscious experience make some sense of an existing issue surrounding possession of phenomenal concepts. Some have claimed that one cannot possess a phenomenal concept unless one has had experiences of the appropriate type (e.g. Papineau (1998): 5). According to this view, if you have not experienced red, you cannot possess the concept RED_{ph}. Call this “the experiential requirement” on possession of a phenomenal concept. This view has the unpalatable implication that blind people, or even sighted people who have experience orange and yellow but not red cannot think the thought STRAWBERRIES CAUSE RED_{PH} EXPERIENCES. The experiential requirement denies the possibility of acquiring a phenomenal concept through interaction with other agents, perhaps with the aid of deference. I have explained elsewhere (Rabin (2011a), Chapter 3: section 6) my objections to the experiential requirement.

There is a good idea in the vicinity of the “possession requires experience” thesis. The possessor of the concept needn’t have experienced the phenomenal quality the concept denotes. She can acquire the concept from someone else. But somewhere down the line of concept acquisition, someone must have experienced that phenomenal quality or one like it.³⁸ When an agent experiences a given phenomenal quality, she is able to demonstrate it and demonstratively apply a (potentially new) phenomenal concept to it. This demonstrative application

³⁸The qualification “or one like it” is intended to handle “missing shade of blue” type cases. An agent could create a concept for a quality she anticipates the existence of, while having experienced only qualitatively similar qualities (e.g. the shades of blue on either side of the missing shade). Or an agent might create a phenomenal concept for a conjunctive quality (purple and smells-like-sulfur) for which she has experienced only the conjuncts.

plays a vital role in giving the phenomenal concept its representational features, including its reference.

These points about phenomenal concepts highlight why even the god-language must rely on demonstrative application, and not just recipes, in the generation of meaning for phenomenal concepts and terms. The gods who have never had phenomenal experiences of, say, color, can't come up with phenomenal color concepts. Once the gods have experienced color phenomenal qualities, they'll demonstrate those qualities as part of the process of creating a concept (or term) for those qualities.

Two features of the process of creating a phenomenal concept are worth mentioning. First, the agent must demonstrate the phenomenal quality. Second, she must demonstrate the quality in the right way. (I won't offer much to clarify "the right way"). Demonstrating by declaring oneself as referring to "Robert Zimmerman's favorite color" is an example of the wrong way. Having the experience in question enables one to demonstrate the quality in the right way. That is why experience of a phenomenal quality is necessary for creating a phenomenal concept or term.

2.7.5 Toward A More Moderate Rationalism: the Moderate Rationalist Link

I've argued that the scope of a priori reasoning is not as vast as the strong modal rationalists and proponents of the recipe view have supposed. This might suggest that there are no interesting answers to the supplementation question. Perhaps we should abandon it. However, I'm hesitant to give up either modal rationalism or the supplementation question. The considerations of section 5 in favor of modal rationalism remain.

In this section, I'd like to use the discussion so far to suggest an amendment

to the strong modal rationalist's proposal. I'll suggest a weaker form of modal rationalism that offers an answer to the supplementation question that is less minimal than the strong rationalist's response but remains strong enough to be interesting. Importantly, this moderate form of modal rationalism won't sanction the anti-physicalist premise (P2) of the master argument.

The trench view stresses the importance of demonstrative applications in the determination of meaning. The reference of 'water' is partly determined by what kind of stuff we happen to have applied 'water' to in the past. Thus, to know what 'water' refers to, to know that water is H₂O, and to know that water covers 70% of the Earth's surface, we need to know what stuff we've applied our term 'water' to in the past. This point suggests an amendment to the strong modal rationalist's proposed answer to the supplementation question. We could add a demonstrative truth to the a priori implication base.

A *demonstrative truth* \mathcal{D} is a set of pairings of representational tokens with the demonstrata to which those tokens were applied. For example, if I once called a particular stream 'water', then the pair $\langle \text{'water'}, \text{this stream} \rangle$ appears in \mathcal{D} . For every time an agent demonstrated some particular(s) and labeled it as an F, the pair $\langle F, \text{the demonstrated particular(s)} \rangle$ appears in \mathcal{D} . The demonstration and/or labeling can occur in language, thought, or perception.

The moderate rationalist answer to the supplementation question says that a necessitation base, in order to become an a priori implication base, must be semantically neutral and supplemented with an indexical truth and a demonstrative truth. This base, which we might call $\mathcal{N}, \mathcal{J}, \mathcal{D}$ (\mathcal{N} for necessitation, \mathcal{J} for indexical, \mathcal{D} for demonstrative), a priori implicates all truths. Because the scope of a priori reasoning is not as vast as the strong rationalist presumes, a neces-

situation base needs additional help to become an a priori implication base. \mathcal{D} provides this help. We get the Moderate Modal Rationalist Link.

(Moderate Modal Rationalist Link): A semantically neutral necessitation base, supplemented with an indexical truth and a demonstrative truth, a priori implicates every truth.

The Moderate Rationalist Link remains an interesting form of modal rationalism. Let's call any form of modal rationalism that endorses this link '*Moderate (Modal) Rationalism*'. Moderate Rationalism satisfies many of section 2.5's motivations for modal rationalism. Moderate Modal Rationalism retains a prohibition on brute unexplained metaphysical necessity. All necessary truths can be explained, on a priori grounds, from the necessitation base combined with the indexical and demonstrative truths. Like Strong Modal Rationalism, Moderate Rationalism retains the spirit of the view from above and the factorization view of the necessary a priori. Moderate rationalists disagree with strong rationalists, claiming the idealized a priori reasoner needs a little more information (provided by the demonstrative truth) before she can figure out how to describe all of modal space. And most importantly, moderate rationalists can claim a strong and important link between metaphysical necessity and a priori justification.

The Moderate Modal Rationalist Link is an ambitious thesis in the philosophy of scrutability. $\mathcal{N}\mathcal{J}\mathcal{D}$ is a sparse set of truths that falls far short of all truths. It would be an interesting result and a vindication of the rationalist programme if all truths could be known on the basis of $\mathcal{N}\mathcal{J}\mathcal{D}$.

2.7.6 Objection: The Hard Problem of Cats

According to the demonstrative reply, the epistemic gap between the physical truths (supplemented with the indexical truth \mathcal{J}) and the experiential truths is at least partly result of the meta-semantic importance of demonstratives in generating the reference of phenomenal terms such as 'red_{ph}', 'blue_{ph}', and 'ticklish_{ph}'. But I've claimed that demonstrative applications play a vital meta-semantic role for non-phenomenal terms as well, including 'cat', 'water', and 'Hesperus'. My story seems to suggest that the importance of demonstratives in generating the reference of 'cat', 'water', and 'Hesperus', along with the absence of demonstratives in the indexically supplemented physical necessitation base, should lead to the result that there is also a lack of a priori implication, and an epistemic gap, between the physical truths and the cat-truths, the water-truths, and the Hesperus-truths. But there's a hard problem of consciousness, and no hard problem of cats. I seem to make the wrong prediction. What gives?

I make several points in response. First, it's not clear that the "hard problem of consciousness" should be spelled out as a claim about a priori implication. There are lots of a posteriori difficulties associated with the scientific study of consciousness and with placing consciousness in our scientific understanding of the natural world. It's doubtful that all the issues associated with "the hard problem" ultimately trace back to a lack of a priori implication between physics and consciousness.

Second, I think it's not so clear that there is a priori implication from the physical truths to the cat-truths. I think the absence of demonstratives places serious obstacles to a priori implication of the cat-truths. Without a demonstrative truth \mathcal{D} to tell us what types of creatures we've been labeling 'cat', we may very well

be unable to use a priori reasoning from $\mathcal{P}\mathcal{J}$ to learn all the truths about cats.

Third, cats are less mysterious than consciousness for reasons that have nothing to do with a priori implication. We know a lot, a posteriori, about cats and their properties. We can see how cats and their properties could arise from, and be metaphysically generated from, the behavior of physical fundamentalia. But we can't see, not even in rough sketch, how consciousness could similarly be generated from the physical. This is an a posteriori fact about our current epistemic situation - a fact that incorporates all the a posteriori psychology, neuroscience, and philosophy that has been done up until this point.

Fourth, and most importantly, the considerations of section 2.7.3 ("The Ineliminable Connection Between Phenomenal Concepts and Demonstratives"), on the ineliminability of the connection between demonstratives and representations for consciousness, offers the disanalogy between the problem of consciousness and the problem of cats that the objection seeks. The connection between our representations of consciousness and demonstrative application to experience is ineliminable. There's no way to get on to consciousness without demonstrating it. In contrast, some agent could, perhaps using the god-language and the term 'cat^g', get on to the cats without using demonstratives (or without relying on them in as heavily a manner as one must for consciousness). The particular hardness of the problem of consciousness stems, at least in part, from this vital and ineliminable connection between consciousness and demonstration of it.

2.8 Conclusion

The metaphysical difficulties surrounding the hard problem of consciousness revolve around a tension between three claims.

(Physicalism) All the fundamental constituents of reality are physical.

(Epistemic Gap) There is an epistemic gap between the physical and conscious experience.

(Modal Rationalism) There are important connections between metaphysical modality and a priori justification.

Physicalism is a purely metaphysical thesis about the nature of the fundamental constituents of reality. It has modal commitments as well. The Epistemic Gap, on the other hand, is soundly epistemic and representational. Modal Rationalism gives this epistemic thesis metaphysical bite. One can resolve the tension by denying physicalism (Chalmers (1996a)) or by denying the epistemic gap (Jackson (2005), Braddon-Mitchell (2003)). But the easiest and most popular response to the master argument is to give up the connection between the epistemic and the metaphysical that the rationalist posits. There are many ways to do this. One could reject inferences from conceivability to possibility or reject the search for the fundamental constituents of reality via the search for a set of truths on the basis of which all other truths can be known a priori.

However, I believe we should not be so quick to reject modal rationalism. I have already laid out my reasons. If we can, we should preserve rationalist connections between the epistemic and the metaphysical, between explanation and fundamentality, and between modality and the a priori. The Demonstrative Reply offers an appealing and well-motivated response to the Master Argument Against Physicalism and the trilemma above. It is compatible with ambitious and substantive forms of modal rationalism. The Demonstrative Reply should appeal to the rationalist physicalists, a group of which I proudly count myself a member.

Chapter 3

Conceptual Mastery and the Knowledge Argument*

Chapter Abstract

According to Frank Jackson's famous knowledge argument, Mary, a brilliant neuroscientist raised in a black and white room and bestowed with complete physical knowledge, cannot know certain truths about phenomenal experience. This claim about knowledge, in turn, implies that physicalism is false. I argue that the knowledge argument founders on a dilemma. Either (i) Mary cannot know the relevant experiential truths because of trivial obstacles that have no bearing on the truth of physicalism or (ii) once the obstacles have been removed, Mary can know the relevant truths. If we give Mary the epistemological capabilities necessary to draw metaphysical conclusions about physicalism, she will, while trapped in the black and white room, be able to know every truth about phenomenal experience.

*This chapter is a revised and improved version of my article "Conceptual Mastery and the Knowledge Argument", *Philosophical Studies*, 54, 125-147, 2011.

3.1 Introduction[†]

According to Frank Jackson's famous knowledge argument (Jackson (1982, 1986)), Mary, a brilliant neuroscientist raised in a black and white room and bestowed with complete physical knowledge, cannot know certain truths about phenomenal experience. For example, she can't know what it's like to see red. This claim about knowledge, in turn, implies that physicalism is false. I argue that the knowledge argument founders on a dilemma. Either (i) Mary cannot know the relevant experiential truths because of trivial obstacles that have no bearing on the truth of physicalism or (ii) once the obstacles have been removed, Mary can know the relevant truths. If we give Mary the epistemological capabilities necessary to draw metaphysical conclusions about physicalism, she will, while trapped in the black and white room, be able to know every truth about phenomenal experience.¹

3.2 Physicalism

Physicalism requires, at the least, that the physical state of the world metaphysically necessitates the complete state of the world. In other words, if you copy the actual world in all its physical respects, you copy the world in all its respects. The knowledge argument purports to refute this modal claim. We operationalize physicalism as follows:

[†]Thanks to Torin Alter, Ned Block, Bill Brewer, Tyler Burge, David Chalmers, Mark Greenberg, Brian Rabern, Jonathan Simon, the editors at *Philosophical Studies*, and audiences at New York University, the University of Warwick, and UCLA for helpful comments and criticism.

¹Phenomenal experiences are associated with raw experiences or sensations. Some examples include the sensation associated with a pain, the color-experience associated with seeing the sky (a "blue experience"), or the sensation one has when the back of one's neck itches. In the words of Nagel (1974), there is something "it is like" for the subject of a phenomenal experience. I use the terms 'phenomenal' and 'experiential' interchangeably.

Physicalism =_{def} Every austere physical duplicate of the actual world is a duplicate simpliciter.²

We duplicate worlds using propositions - if all the same propositions are true at two worlds, they are duplicates.³ An *austere physical duplicate* of the actual world is a metaphysically possible world at which every austere physical proposition true at the actual world is true. A duplicate simpliciter of the actual world is a world at which every proposition true at the actual world is true. An *austere physical proposition* (or *physical proposition*) is a proposition composed of concepts taken only from physics, mathematics, and logic, or which is expressible using only vocabulary taken from physics, mathematics, and logic.⁴ A *truth* or *fact* is a true proposition. What this definition of physicalism says is this: any possible world with the same austere physical state as the actual world, i.e.

²In all likelihood, this is not a good *definition* of physicalism. I prefer a definition in terms of the notion of *fundamentality* (cf. Chapter 1). But even if one does not think this a good definition of physicalism, the modal claim embodied therein is almost certainly a necessary condition for physicalism. The modal claim is a minimal commitment shared by many forms of physicalism. The knowledge argument attacks physicalism via this necessary condition. Thus it serves our purposes here to utilize this operationalized definition.

³Using propositions to account for when two worlds are the same or different requires one substantive assumption - that any way a world is can be expressed by some proposition. All states of affairs can be expressed propositionally. This assumption is trivial if we treat propositions as sets of worlds. I do not treat them as such. I treat propositions in a broadly Fregean manner, as structured entities composed of concepts. On this understanding of a proposition, there are some reasons to doubt the assumption, including the possibility of inexpressible states of affairs. However, such fringe cases are irrelevant to the main narrative here. I will ignore them and treat the assumption as a given.

⁴On this terminology, the propositions expressed by ' $2+2 = 4$ ' and 'predicate logic is complete' are austere physical. This is slightly odd. I think it helpful to think of the "austere physical" propositions as those propositions which are *physicalistically respectable*. In the debate between the physicalist and dualist, it is certainly legitimate for the physicalist to appeal to mathematics and logic. The physicalist claims that our world is, at the fundamental level, purely physical. But this does not mean that the fundamental layout can be expressed without using any math or logic. The austere physical propositions are those to which the physicalist may legitimately appeal to in describing the world's fundamental layer (or necessitation base) without jeopardizing his or her materialist credentials.

with the same arrangement of quarks, leptons, and bosons, same distribution of physical properties (mass, charge, spin), and same physical laws, will have the same complete state as the actual world, including the same distribution of macroscopic objects and mental states.

I assume a fine-grained, roughly Fregean treatment of propositions.⁵ On this approach, the proposition expressed by ‘Hesperus is bright’ is not the same proposition expressed by ‘Phosphorous is bright’. Concepts compose propositions in a manner similar to that by which words compose sentences. I will use small capitals to express concepts. ‘HESPERUS’ denotes the Hesperus-concept. On the Fregean treatment, the propositions HESPERUS IS BRIGHT and PHOSPHOROUS IS BRIGHT are not the same proposition.

3.3 The Knowledge Argument

Here’s our quick and dirty starting version of the knowledge argument.

- (1) Knowledge of all the austere physical information does not put one in a position to know all the information.
- (2) If (1), then physicalism is false.
- (3) Therefore, physicalism is false.

The argument is clearly valid. But is it sound? Neither premise is obvious;

⁵This assumption is not uncontroversial, but it is dialectically fair. It *helps* the knowledge argument, which I plan to argue against. The knowledge argument requires there to be propositions that Mary cannot know. The Fregean approach individuates propositions more finely than some other approaches (e.g. Russellian propositions). If Mary cannot know some coarse-grained proposition, it follows that she can’t know some more fine-grained proposition. The converse implication does not hold. Thus, using a fine-grained approach makes it easier for the knowledge argument to achieve its goal of demonstrating the existence of propositions Mary cannot know. By choosing a fine-grained approach to propositions, I have chosen to battle the knowledge argument on its preferred terrain. For example, a fine-grained approach to propositions makes it much more difficult to launch the “same proposition, different mode of presentation” defense against the knowledge argument (cf. Byrne (2002), Tye (1995)). I’ll argue that, even given a fine-grained conception of propositions, Mary can know the relevant truths.

both have been coherently denied by physicalists.⁶ Support for premise (2) comes from theorizing about modality. Support for premise (1) comes from Jackson (1982, 1986)'s famous thought experiment involving Mary.

Jackson asks us to envision Mary, a "brilliant neuroscientist who is... forced to investigate the world from a black and white room via a black and white television monitor. She... acquires... all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on." Mary has complete physical information, i.e. she knows every austere physical proposition, and every proposition that such knowledge places one in a position to know. Jackson asks whether Mary learns anything new when she emerges from the black and white room and sees a ripe red tomato for the first time. He concludes that "it just seems obvious that she will learn something about the world and our visual experience of it." Therefore, knowledge of all the austere physical information does not put one in a position to know all the information - i.e. premise (1) is true.⁷

Premise (2) gets support from the desire to avoid mysterious unexplained brute necessities and from *modal rationalism*, according to which there are important (potentially constitutive) connections between modality and epistemological notions like "in a position to know", "a priori", and "conceptual truth".⁸ The general physicalist idea that "all information is physical information", knowable on the basis of complete austere physical information, also supports premise (2).

⁶Premise (1) has been denied by so-called "type-A" physicalists, premise (2) by so-called "type-B" physicalists. This terminology originates in Chalmers (2003a).

⁷Enthymematic here is the claim that if Mary can't know, no one can.

⁸Jackson (1998) is the most sustained argument to date for modal rationalism. Cf. also Chalmers (forthcoming, 1999), Chalmers & Jackson (2001), Levine & Trogon (2009), and Schwarz (2007).

The knowledge argument starts from premises about epistemology (what Mary can't know) and draws a metaphysical conclusion: the falsity of physicalism. Physicalism is a claim about how the world is, not a claim about how we know about the world. From the start, it should strike the reader as a bit mysterious how premises about knowledge allow one to draw conclusions about metaphysics. Unfortunately, I won't do much to alleviate the mystery here. There are loads of interesting issues in this area; I will not even begin to explore them.⁹ Instead, my strategy will be to provisionally accept the general idea that one can, in general, draw metaphysical conclusions from epistemology in the manner that the knowledge argument does. I'll argue that one can grant this much and still resist the knowledge argument.¹⁰

The knowledge argument relies on a specific type of inference. The inference moves from an epistemic gap, the lack of knowability of a *target set* of propositions on the basis of a *base set*, to a metaphysical gap, the lack of necessitation from the base set to the target proposition(s).¹¹ In the knowledge argument, the

⁹Interested readers can consult Chalmers (1999, 2012), Jackson (1998), and my dissertation.

¹⁰Replying to the knowledge argument in this way, by provisionally granting that one can get metaphysical bite from epistemological worries, offers significant dialectical advantages. One could respond to the knowledge argument with the dismissive reply. *The dismissive reply* rejects *tout court* the idea that one can make knowledge argument style inferences from epistemology to metaphysics. But there are many reasons to accept such inferences. The link between conceivability and possibility provides one example of an epistemology-to-metaphysics inference. Those philosophers who want to preserve some form of a rationalist link between epistemology and metaphysics can't respond to the knowledge argument using the dismissive reply. They need a more sophisticated reply. This paper offers such an option.

¹¹A question that may have occurred to the reader is whether we are *ever* justified in inferring from the lack of knowability of a target proposition from a base set to the fact that the base does not necessitate the target. This issue becomes more pressing when one recognizes the obvious counterexamples. Even if one possesses all the relevant concepts, knowing that there is water in the glass does not put one in a position to know that there is H₂O in the glass. Yet the fact that there is water in the glass does necessitate that there is H₂O in the glass.

The worry can be addressed and the obvious counterexamples avoided. The modal rationalist advocates of the general strategy of inferring from an epistemic to a metaphysical gap (e.g. Jackson (1998), Chalmers (1996b), Chalmers & Jackson (2001)) do not endorse the general claim that,

base is the set \mathcal{P} of all true austere physical propositions. The target set contains, at the minimum, one proposition E about phenomenal experience (perhaps a proposition about what it's like to see red). Because Mary knows all the propositions in \mathcal{P} , but is not thereby in a position to know E , \mathcal{P} does not necessitate E . Since \mathcal{P} does not necessitate E , physicalism is false. Or at least so the knowledge argument maintains.¹²

3.4 Roadmap

We'll see shortly (section 5 - "The Missing Concept Reply") that the argument can't be as simple as I've just described. If metaphysical conclusions are to be drawn from the epistemic gap that Mary's case indicates, some epistemological issues must be finessed first. Much of the paper will be spent finessing those issues. However, before diving in, it will be helpful to frame the dialectic and outline the moves to come. The knowledge argument starts from the lack of an epistemic relationship between the set of physical propositions \mathcal{P} and the target

for any propositions B and T , if knowing B does not put one in a position to know T , B does not necessitate T .

The inference to a metaphysical gap is valid only for a special class of base propositions. There is insufficient space for a full treatment here. One strategy is to require that, in order to draw conclusions about necessitation from facts about epistemic implication, the base propositions must be *semantically neutral* - i.e. not "twin-earthable". Famously, Putnam (1975) showed that 'water' is not semantically neutral. It is plausible that the vocabulary of fundamental physics, the vocabulary in which the austere physical description \mathcal{P} is written, is semantically neutral. Anything that acts like an electron *is* an electron, whereas something can act like water and be XYZ, not water. Even if the fundamental vocabulary of physics is not semantically neutral, this merely introduces an epicycle on the dialectic (cf. Chalmers (1996b): 135-6). Another important difference between the proposition that there is water in the glass and the proposition \mathcal{P} under consideration in the knowledge argument is that, according to the physicalist, \mathcal{P} necessitates all the facts - it leaves the truth value of no proposition unsettled. The proposition that there is water in the glass has no such pretensions.

¹²There are other ways to construe the knowledge argument. I do not consider them all here. I deal only with the knowledge argument construed as an attack on the modal commitments of physicalism. These commitments are represented in the operationalized definition of physicalism from page 2, which claims that the austere physical state of the world necessitates the complete state. As I construe it, the knowledge argument challenges this claim.

experiential propositions. This epistemic relationship must meet two desiderata. First, its absence must be good evidence for a lack of necessitation. Second, this epistemological relationship must fail to hold between the austere physical propositions \mathcal{P} and certain experiential propositions. Evidence for the failure of this epistemological relation comes from considering what Mary can and cannot know. She is our litmus test. I introduce a technical term, *implication* (verb form: *implicate*), to express an epistemic relationship that might meet the desiderata. 'Implicate' allows us to express the following schema for the knowledge argument:

- (1) \mathcal{P} does not implicate some proposition E about phenomenal experience.
- (2) If \mathcal{P} does not implicate E, then \mathcal{P} does not necessitate E, and physicalism is false.
- (3) Therefore, physicalism is false.

Substituting for 'implicate' here yields different versions of the knowledge argument. A candidate notion of implication meets the two desiderata above if and only if it validates both premises. One upshot of this paper is that finding such a conception of epistemological implication is difficult. The rest of the paper will be spent considering three different candidate notions of implication. All such candidates founder on one premise or the other. The final result is that the knowledge argument fails. In the next section I consider the most obvious notion of epistemological implication:

- **(Implication = in a position to know)** B implicates T iff any agent who knows all the propositions in B is thereby in a position to know T. ('B' stands for *base*, 'T' for *target*).

I'll argue that this conception of implication fails to vindicate the second premise of the knowledge argument schema. Inferring a lack of necessitation from a failure of implication in the "in a position to know" sense is a mistake. The basic reason is that the agent might know the propositions in B but not possess the concepts composing the propositions in T and thereby be unable to know T because the agent cannot even consider T. But this type of failure shows nothing about necessitation (see p. 7-8).

In section 6 ("The Conceptual Mastery Reply"), I consider a more sophisticated version of implication:

- **(Implication = in a position to know given concept possession)** B implicates T iff any agent who (i) knows all the propositions in B and (ii) possesses all the concepts in T is thereby in a position to know T.

I'll argue that this conception of implication also fails to vindicate the second premise, and does so for reasons largely similar to the previous conception.

The final, and most sophisticated conception of implication goes as follows:

- **(Implication = in a position to know given conceptual mastery)** B implicates T iff any agent who (i) knows all the propositions in B and (ii) has mastered all the concepts in T is thereby in a position to know T.

For this conception of implication to make sense to the reader, I'll need to distinguish between two relations one can have to a concept. One can possess a concept or one can master a concept. I explain the difference in section 6. In sections 8-9, I'll argue that this conception of implication fails to vindicate the first premise (unlike the first two conceptions, which failed to vindicate the second premise). In other words, once Mary has mastered the concepts in the experiential proposition E, she will, given knowledge of all the austere physical propositions in

\mathcal{P} , be able to know E. In effect, I claim that once we target the appropriate epistemological relationship - “in a position to know, given conceptual mastery” - the basic claim of the knowledge argument fails. Despite being raised in the black and white room, Mary will be in a position to know every truth about conscious phenomenal experience.

In the end, I propose that the knowledge argument founders on the following dilemma:

(The Dilemma) Either Mary does not have conceptual mastery of experiential concepts, in which case we cannot draw any metaphysical conclusions on the basis of what she can and cannot know, or Mary does have conceptual mastery, and, with the help of knowledge allotted to her by the informational assumption, she will be able to know all the relevant experiential propositions.

The informational assumption says that if some proposition is necessitated by \mathcal{P} , then Mary can know it. Proponents of the knowledge argument need the informational assumption (the argument’s second premise is an instance of the assumption’s contrapositive). I plan to wield it against them. In section 8 (“The Informational Assumption and the Second Horn”), I’ll discuss in detail both this dilemma and the informational assumption.

3.5 The Missing Concept Reply

Consider the following version of the knowledge argument (copied from above):

- (1) Knowledge of all the austere physical information does not put one in a position to know all the information.
- (2) If (1), then physicalism is false.
- (3) Therefore, physicalism is false.

The knowledge argument, as stated, does not stand a chance. It is a mistake to conclude that a base set of propositions does not necessitate a target proposition from the fact that someone who knows the base propositions is not in a position to know the target. An agent might know the base propositions but not even *possess* the concepts in the target proposition. As a result, the agent won't be in a position to know the target propositions - because he or she can't even consider them! But an epistemic gap of this nature shows *nothing* about necessitation. The following example demonstrates why. Imagine that Mary's sister Jane knows all the information in \mathcal{P} , all the information about the state of Ralph's body, including where he does and does not have inflammation. But Jane does not possess, and is not in a position to possess, the concept ARTHRITIS. Jane knows that Ralph does not have rheumatoid inflammation in any of his joints. Ralph has aches and pains, but no arthritis. But Jane does not know, and is not in a position to know, that Ralph does not have arthritis. She can't know because she does not possess the concept ARTHRITIS. If we infer from this epistemological gap to a metaphysical gap, we'll conclude that Ralph's lacking inflammation of the joints does not necessitate that he does not have arthritis. But clearly this is false. Arthritis just is rheumatoid inflammation of the joints. If Ralph lacks inflammation in his joints, he thereby lacks arthritis. Ralph's lacking inflammation in the joints *does* necessitate that he does not have arthritis.

The lesson is that if we're going to infer from an epistemic gap between two propositions (or sets of propositions) to a lack of necessitation, and we test for the epistemic gap by checking what hypothetical agents could come to know, we must require that these agents possess the concepts in the target proposition, or at

least are in a position to possess them.¹³ The fact that Jane, who does not possess the concept ARTHRITIS, can't know that Ralph lacks arthritis despite knowing that he has no joint inflammation, shows nothing about a lack of necessitation between the propositions RALPH HAS NO JOINT INFLAMMATION and RALPH DOES NOT HAVE ARTHRITIS.

In the case of Mary, it is plausible that Mary does not possess the relevant phenomenal concepts.¹⁴ For example, consider the concept RED_{exp}. RED_{exp} is the phenomenal, experiential, concept of red. RED_{exp} applies to experiences with a certain qualitative character (you know the one). Neither light-waves nor objects can be red_{exp}. Tomatoes, fire trucks, and strawberries are red, but they cause red_{exp} sensations.¹⁵ The English term 'red' is likely polysemous between, or a mongrel of, the red-concepts of light waves, of surfaces, and of sensations. I introduce a new term, 'red_{exp}', and concept RED_{exp}, to lock on to the type of red that only sensations can instantiate.

The fact that Mary has never had a red_{exp} experience, cannot imagine a red_{exp} sensation, and could not recognize a red_{exp} sensation as red_{exp} were she to have one, supports the claim that Mary does not possess the concept RED_{exp}. With the

¹³One might object that Jane is at least in a position to possess the concept ARTHRITIS - all she need do is combine the concepts RHEUMATOID, INFLAMMATION, OF, THE, and JOINTS. In reply, I make three points. First, it's far from clear that Mary does possess those concepts. She clearly has all the austere physical concepts. But it's highly doubtful that the concept INFLAMMATION can be generated by combination from austere physical concepts like QUARK and MASS. Second, it's not clear that ARTHRITIS is merely a composite of those other concepts. Lacking inflammation of the joints can necessitate lacking arthritis without ARTHRITIS being a composite of the concepts INFLAMMATION, OF, THE, and JOINTS. Third point: Suppose that Mary is in a position to possess ARTHRITIS. It still remains true that were Mary not in a position to possess ARTHRITIS (perhaps *per impossibilia*), then one should not infer from her inability to know that Ralph lacks arthritis to the claim that Ralph's lack of joint inflammation does not necessitate that he lacks arthritis. Thus even if the objection holds, the lesson remains: don't infer lack of necessitation from a case in which the epistemic agents is not in a position to possess concepts in the target propositions.

¹⁴I use the terms 'experiential concept' and 'phenomenal concept' interchangeably.

¹⁵All tomatoes mentioned in this paper are ripe tomatoes with brilliant red skins.

“concept possession requirement” in place, the physicalist has an easy reply to the knowledge argument - deny premise (2). It is foolish to conclude, from the fact that Mary’s knowledge of base physical propositions does not put her in a position to know target experiential propositions, that the base does not necessitate the target. Mary can’t even *consider* the target propositions because she does not possess the concepts that compose them! Only an inability to know, for an epistemic agent that possesses (or is in a position to possess) all the concepts in the target propositions, has any hope of yielding the desired metaphysical consequences.¹⁶

This discussion yields an obvious suggestion for how to fix the knowledge argument. Perhaps “B implicates T” means that any agent who has knowledge of B, and possesses all the concepts in T, is thereby in a position to know T. In the next section I argue that this suggestion fails.

3.6 The Conceptual Mastery Reply

Some philosophers have pointed out that Mary, raised in her black and white room, *can* possess the relevant experiential concepts.¹⁷ She can acquire them from her co-workers, who have seen ripe red tomatoes and the big blue sky. Mary won’t have as good a grasp of these concepts as her co-workers. After all, Mary has never experienced red_{exp}. But Mary will at least possess the relevant concepts and be able to grasp propositions that contain them.¹⁸ However, it seems that

¹⁶Byrne (2006), Hellie (2004):348, ? :474-5, and Tye (2000):17-8 all make similar points. They all note that one cannot draw any substantive conclusion from the Mary case if Mary does not possess experiential concepts and cannot even consider the experiential propositions in question.

¹⁷Ball (2009) and Tye (2009), among others, make this point.

¹⁸A full defense of the claim that Mary can possess the relevant concepts in this way would take us too far afield. It is certainly the consensus view in philosophical research on content. Here’s a brief argument for the claim. If Mary thinks that the sky causes red_{exp} sensations, her thought is

Mary's acquisition of these concepts from her co-workers won't help her learn the missing experiential information. If so, then Mary can possess all the concepts in the target experiential propositions, know all the austere physical propositions, and yet still not be in a position to know the experiential propositions. The knowledge argument can overcome the objection raised in the previous section and adapt to incorporate its lesson, which is that the relevant epistemic agents must possess the concepts in the target proposition.

In this section, I argue that we need additional constraints on implication. We should not infer from "knowledge that B does not put someone who possesses all the concepts in T in a position to know that T" to the fact that B does not necessitate T. Possessing the concepts in T is not enough; we must require more of the epistemic agent.

There are (at least) two different levels of understanding a concept. The first is *concept possession*. If one is able to grasp propositions that contain the concept, or think propositions a component of which is the concept, then one possesses the concept. Concept possession is easy to obtain. Burge (1979)'s Alfred possesses the concept ARTHRITIS, but Alfred does not know what every doctor knows - that arthritis affects joints and not limbs. Some of the doctors have a more sophisticated and demanding relation to the concept ARTHRITIS - they have *conceptual mastery*. The proverbial "experts", to which laymen defer, often possess

wrong - the sky causes blue_{exp} sensations, not red_{exp} ones. This is so because her thought has the same truth-conditions the thought we would express with 'the sky causes red_{exp} sensations' does. It has the same truth conditions because it contains the same concepts, which contribute to the truth-conditions of the complete thought. Mary's concept has the same veridicality conditions our concept RED_{exp} does. Therefore, Mary thinks with the same RED_{exp} concept the rest of us do.

Here's another brief argument: If Mary does not think with our RED_{exp} concept, she thinks with some other idiosyncratic concept (perhaps her RED_{exp} concept is similar to our BLUE_{exp} concept). Interpreting her charitably, her idiosyncratic thought is most likely correct. But her thought is not correct; she is wrong. Therefore she does not think with another concept.

conceptual mastery. One can possess a concept and still be grossly wrong about its extension, and even about constitutive a priori truths regarding the concept. Conceptual mastery is less tolerant of such errors.¹⁹

Once we have distinguished these two levels of understanding a concept, we can improve our grip on the connection between epistemic gaps and absence of necessitation. We said above that when inferring from lack of implication to lack of necessitation, only lack of knowability for agents who possess all concepts in the target proposition is relevant. We can do better: only lack of knowability for agents who have mastery of all concepts in the target proposition is relevant. The following example demonstrates why.

Imagine again Mary's sister Jane. By engaging with her co-workers, Jane comes to possess the concept ARTHRITIS. She knows that Esther, who has a rheumatoid inflammation in her knee, has arthritis. She knows that Hilary does not have arthritis. But Jane is not a master of the concept ARTHRITIS; she thinks that one can have arthritis in the forearm. Jane has complete knowledge of \mathcal{P} and knowledge of who has what type of inflammation where. She knows that Ralph does not have any ailments of the joints, but he does have inflammation in his forearm. Unlike the previous case, because Jane now possesses the concept ARTHRITIS, she can consider the proposition that Ralph does not have arthritis.

¹⁹The distinction between concept possession and conceptual mastery also appears in Greenberg (MS, 2009) and Bealer (2002b). Of course, the distinction has its roots in the seminal discussions of Burge (1979) and Burge (1986). Mark Greenberg deserves credit for stressing to me the importance of the distinction and the pitfalls into which philosophers can fall when they ignore or pay insufficient attention to it. In several places (Ibid.), Greenberg uses an argumentative strategy similar to the one I adopt here. His use pre-dates mine. The strategy involves identifying a philosophical argument or position that implicitly assumes that possession of a concept entails mastery. (This assumption is often masked by an ambiguity, between possession and mastery, in the terminology of "understanding / possessing a concept"). This assumption undermines the position. Once we restrict attention to cases in which the assumption holds, the argument fails to accomplish its goal. For example, I will argue (sections 8-9) that once we focus on cases in which Mary has conceptual mastery, she can know the target experiential propositions.

But because of her erroneous view that one can have arthritis in non-joints, Jane will not come to know that Ralph does not have arthritis. If we don't require, in the inference from an epistemic to a metaphysical gap, that the relevant epistemic agents have conceptual mastery (i.e. if mere concept possession is enough), then we will be forced to conclude, on the basis of Jane's inability to know that Ralph does not have arthritis, that Ralph's lacking any ailment of the joints does not necessitate that he does not have arthritis. But that conclusion is preposterous. Thus, we must require, if we are to infer from a lack of implication to a corresponding lack of necessitation, that the epistemic agents have mastery of all concepts in the target proposition(s).²⁰

The conceptual mastery requirement offers a reply to the amended version of the knowledge argument in which Mary does possess experiential concepts such as RED_{exp}.²¹ Mary, despite coming to possess the concept RED_{exp} via interaction with her co-workers, does not have conceptual mastery. In order to have conceptual mastery of RED_{exp}, one must be able to identify a red_{exp} sensation when experiencing one. One might also maintain that conceptual mastery of RED_{exp} requires that one have experienced red_{exp} and/or be able to imagine a red_{exp} patch. Mary has never had a red_{exp} experience, is unable to imagine a red_{exp} sensation, and cannot identify a red_{exp} experience as an instance of her concept RED_{exp}. Therefore, she does not have conceptual mastery of RED_{exp}. As a result, Mary's inability to know propositions containing RED_{exp} on the basis of the austere physical information shows nothing about necessitation. Only if such an

²⁰For obvious reasons, the epistemic agent must have conceptual mastery of all concepts in the base propositions as well.

²¹The missing concept reply has appeared in the literature before (e.g. Hellie (2004), Loar (1990), ?, and Tye (2000)). To the best of my knowledge, the "conceptual mastery" reply has not.

epistemic gap occurs for an agent with mastery of RED_{exp} can we begin to make claims about necessitation.²²

3.7 Aside: The Phenomenal Concept Strategy

The philosophical study of conscious experience and the way we think about it has fostered a large body of literature on *phenomenal concepts*. Phenomenal concepts characterize the nature of conscious phenomenal experience by the way those experiences feel to the subject. RED_{exp} is one such concept. In this section, I take a moment to highlight some applications of the points made thus far to the so-called “phenomenal concept strategy” and to phenomenal concepts in general. The phenomenal concept strategy responds to the knowledge argument and its siblings (such as the zombie argument) by appealing to special features of phenomenal concepts to explain how a lack of epistemic implication between the physical and the experiential truths could be compatible with physicalism.²³ The strategy locates the problem in our ways of thinking about consciousness rather than in the metaphysics of consciousness.

In section 5, we considered one implementation of the phenomenal concept strategy, “the missing concept reply”, which uses facts about the conditions under which a thinker can possess a phenomenal concept to defend against the knowledge argument. In this section, I will argue that fans of the phenomenal concept strategy have misplaced their allegiance. They have focused too much on concept possession and not enough on conceptual mastery. Switching to concep-

²²Thus far, I’ve argued that if we are to draw conclusions about necessitation, we must require that the epistemic agents we use to check for an epistemic gap must both possess and have conceptual mastery of the relevant concepts. I will not argue the point here, but I actually think that even more is required. Cf. Rabin (MS).

²³Cf. Chalmers (2003a): 107 for the zombie argument.

tual mastery yields a vastly improved version of the phenomenal concept strategy. The conceptual mastery oriented incarnation of the phenomenal concept strategy I suggest is more powerful, less vulnerable, and more plausible than previous versions.

In a strict sense, this section is not part of paper's main argument. But for many readers, especially those familiar with the vast literature on the phenomenal concept strategy (by far the dominant response to the knowledge argument), this section should prove interesting enough to merit its inclusion. The impatient reader can skip to the next section. For later argumentation, the only essential information from this section is the introduction of the three candidate requirements on possession and/or mastery of a phenomenal concept (which can be found immediately below).

We can delineate three different requirements on possession of an experiential/phenomenal concept.

- *Experiential requirement*: Possession of a phenomenal concept requires having had experiences that fall under the concept.
- *Imaginative requirement*: Possession of a phenomenal concept requires the ability to imagine instances of the experiential quality that falls under the concept.
- *Recognitional requirement*: Possession of a phenomenal concept requires the ability to recognize experiences that fall under the concept as falling under the concept.

Fans of the phenomenal concept strategy have endorsed versions of these re-

quirements, and others still.²⁴ I deny them all; concept possession is too easy to obtain.

Historically, many defenders of the phenomenal concept strategy have endorsed the experiential requirement on possession of a phenomenal concept. Papineau (1998): 5 writes that anyone “who has never seen anything red cannot deploy a phenomenal concept of red visual experience.” According to this line, Mary has not experienced red_{exp} and does not possess the concept RED_{exp} . Since she cannot consider the relevant propositions, her inability to know them demonstrates nothing about physicalism. This strategy has been popular, but less than effective.

Once the possibility of possessing a concept without mastery, potentially through deference to others, is recognized, the experiential requirement looks implausible. Blind people, who have never seen red_{exp} , can possess the concept RED_{exp} . They can truly think and say, “Tomatoes cause red_{exp} experiences.” Ball (2009) and Tye (2009) correctly press this objection against the experiential requirement and its use against the knowledge argument.²⁵ They point out that one can run a knowledge argument in which Mary does possess the relevant phenomenal concepts, but in which this possession will not help her learn the truths about ex-

²⁴Papineau (1998): 5 endorses the experiential requirement. Loar (1990) endorses the recognitional requirement. For a good discussion cf. ?.

²⁵Ball (2009) and Tye (2009) use the term ‘phenomenal concept’ for any concept possession of which requires having had experiences that fall under that concept. On this terminology, the experiential requirement on possession of a phenomenal concept is true by definition. Ball denies that there are any such concepts. But neither Ball nor Tye denies that there are experiential or phenomenal concepts in my sense - concepts that characterize experiences by the way those experiences feel to the subject. Ball and Tye also use the term ‘phenomenal concept strategy’ for a response to the knowledge argument that endorses the experiential requirement on possession of experiential/phenomenal concepts. I use ‘phenomenal concept strategy’ more generally, applying it to any view that appeals to special features of experiential or phenomenal concepts to explain how necessitation from the physical to the experiential is compatible with a lack of epistemic implication.

perience (this is effectively the version of the argument I suggested at the beginning of section 6). Unfortunately, this move does not cut much ice. As argued in the previous section, the epistemic capabilities of someone who merely possesses the concept indicate nothing about necessitation, and hence cannot be brought to bear on the question of physicalism. Epistemic considerations involving agents with less than conceptual mastery are simply a non-starter.

The experiential, imaginative, and recognitional requirements on concept possession have obvious analogs for conceptual mastery. I do not bother to state them. All three requirements are more plausible for conceptual mastery than for concept possession. One can't draw metaphysical conclusions from absence of epistemic implication unless the epistemic agents have conceptual mastery. Thus, as a reply to the knowledge argument, an experiential requirement on conceptual mastery of experiential/phenomenal concepts can do all the same work as an experiential requirement on possession. But the experiential requirement on conceptual mastery of RED_{exp} is not subject to the Ball/Tye-style criticism from deference. The original experiential requirement (on possession) looks bad because it's very plausible that Mary can possess RED_{exp} without ever having had a red_{exp} experience. She can acquire the concept from others who have had red_{exp} experiences. On the other hand, it's not obvious that Mary can have conceptual mastery of RED_{exp} without having had such an experience. I advise fans of the original experiential requirement to transfer their allegiances to the experiential requirement on conceptual mastery. It can do all the same work in replying to the knowledge argument without incurring the obvious costs.²⁶ Furthermore, once

²⁶Alter (forthcoming) replies to Ball (2009) and Tye (2009) in a similar way, and offers the same advice. He also makes the important observation that the epistemic capabilities of mere possessors of a concept are not good indicators of metaphysical necessitation. However, this observation plays a far less central role in his discussion than in mine. Alter has his own version of the concept

we distinguish between concept possession and conceptual mastery, the conceptual mastery reply is a natural extension of the original phenomenal concept strategy.

I neither endorse nor deny the experiential or the imaginative requirement on conceptual mastery of experiential concepts. The experiential requirement seems counterexample-able by a “swamp case”. Take a conceptual master of RED_{exp}. Imagine an atom-by-atom duplicate, generated by erratic random forces in the swamp mist. It is plausible that the duplicate will have conceptual mastery of RED_{exp} despite never having seen a red thing. Whether one can be a conceptual master of RED_{exp} without satisfying the imaginative requirement seems more questionable. However, one might be willing to attribute conceptual mastery to creatures who lacked certain cognitive imaginative capacities, but nonetheless had experienced red_{exp} and were very reliable in their red_{exp} judgements.

Perhaps the requirements can be softened by including a clause about “normal conditions” - i.e. conditions that allow for exceptions involving erratic swamp forces, genies, and fantastical neuro-surgery. Officially, I endorse only the recognitional requirement on conceptual mastery of experiential concepts.²⁷ How-

possession / conceptual mastery distinction. He uses the terminology of possessing a concept under a “deferential” or “non-deferential” mode of presentation. This terminology is misleading for two reasons. First, on the most natural extension of Fregean modes of presentation to our current framework, concepts *are* modes of presentation. Talk of modes of presentation *of* modes of presentation is misleading at best. Second, whether an individual *defers* to someone else, in the sense of accepting correction regarding the use of his or her concept, is only a defeasible guide to whether he or she has mere concept possession or mastery. One can use the term ‘deferential possession’ as a technical term for “possessing a concept without conceptual mastery, in a manner similar to Burge’s arthritis-man does”. But if deferential possession is *defined* in this way, then there is an open question about whether the fact that an individual defers, in the sense of accepting correction, implies anything about whether he “deferentially possesses” the concept in the defined sense.

²⁷Disclaimer for the counterexample mongers: I don’t endorse the *crazy* version of the recognitional requirement. A conceptual master of BLUE_{exp} need not be infallible in her blue_{exp} judgements. If she is drunk, temporarily mentally impaired, or some such thing, all bets are off.

ever, even without the experiential requirement on conceptual mastery, it remains quite plausible that Mary does not have conceptual mastery of RED_{exp} . Her inability to know experiential propositions cannot be wielded against physicalism.

3.8 The Informational Assumption and the Second Horn

The second premise of the Knowledge Argument requires an inference of the following type: “ \mathcal{P} does not implicate T, therefore \mathcal{P} does not necessitate T.” The contrapositive of this inference is: “ \mathcal{P} necessitates T, therefore \mathcal{P} implicates T.” In this section, we turn our attention to this contrapositive. I lay out the *Informational Assumption*, which says that Mary can know every proposition necessitated by \mathcal{P} . The knowledge argument needs the informational assumption for two reasons. First, the informational assumption provides the link between epistemology and metaphysics on which the knowledge argument relies. Second, not making the informational assumption hands the physicalist an easy reply to the knowledge argument.

Suppose the informational assumption is false. Then there is some proposition R (perhaps the proposition that the rate of inflation in Australia is increasing), necessitated by \mathcal{P} , that Mary cannot know. Suppose also that premise (1) of the knowledge argument is true - despite complete physical knowledge, Mary can't know that tomatoes cause red_{exp} sensations. The dualist wants to conclude that \mathcal{P} does not necessitate that tomatoes cause red_{exp} sensations (they could have caused $blue_{exp}$ sensations instead). But the existence of exceptions to the informational assumption gives the physicalist an immediate reply. She'll say that the proposition that tomatoes cause red_{exp} sensations is, like R, one among many propositions that are necessitated by \mathcal{P} yet not knowable by Mary. To cut off this reply the dualist needs the informational assumption – there are no propositions

like R.

From here on, we make the informational assumption, and assume that Mary, using her knowledge of \mathcal{P} , can know every proposition necessitated by \mathcal{P} .²⁸ With the help of the informational assumption, we precisify the basic dilemma of this paper.

The Dilemma: Either Mary does not have conceptual mastery of experiential concepts, in which case we cannot draw any metaphysical conclusions on the basis of what she can and cannot know, or Mary does have conceptual mastery, and, with the help of knowledge allotted to her by the informational assumption, she will be able to know all the relevant experiential propositions.

The dilemma is related to my claim that it is difficult to find a notion of epistemic implication that satisfies both premises of the knowledge argument schema of section 4 (page 5). On the first horn of the dilemma, Mary lacks conceptual mastery, and we cannot draw metaphysical conclusions from her case. On any notion of implication that does not require epistemic agents to have conceptual

²⁸The dualist can accept some exceptions to the informational assumption without weakening her dialectical position. For example, suppose dualism is true and the experiential truths are not necessitated by the austere physical state of the world described in \mathcal{P} . Let 'Bob' be a name whose reference is fixed by the description 'the actual greenest thing in the universe'. Suppose Bob is, in fact, a meadow in Vermont. It is plausible that the proposition BOB HAS SURFACE AREA OF MORE THAN 10 M.² is necessitated by \mathcal{P} despite the fact that someone who knows \mathcal{P} will not (if dualism is true) be in a position to know it (because he or she will not be in a position to know the experiential facts about green_{exp}). We seem to have a counterexample to the informational assumption. The lesson is that the knowledge argument is compatible with the existence of *some* propositions that are both necessitated by \mathcal{P} and not knowable on the basis of \mathcal{P} .

Advocates of the knowledge argument must hope that there are not too many exceptions to the informational assumption. Every exception opens up space for the physicalist to reply to the knowledge argument with the claim that the proposition that tomatoes cause red_{exp} sensations is also an exception, and hence Mary's inability to know it does not demonstrate a lack of necessitation. Thankfully, the exceptions look to be of a special sort. Many of them will contain rigidified descriptions involving phenomena not necessitated by the physical (e.g. dualistic experiential properties). In this paper, the informational assumption will be brought to bear only on propositions that are not of this special type. If the reader thought that many or most terms (or concepts) were rigidified versions of phenomenal descriptions, then the informational assumption would not be of much use. I assume this is not the case.

mastery, premise (2) of the schema is false. On the second horn, Mary has conceptual mastery. But this mastery, in combination with the knowledge allotted to Mary by the informational assumption, will generate an epistemic route to knowledge of experiential propositions. On this horn, premise (1) of the schema is false - Mary can know.

I have already argued for the first horn of the dilemma. In the next section I argue for the second horn. I demonstrate exactly how Mary's conceptual mastery of, e.g. RED_{exp} , in combination with the informational assumption, yields an epistemic route to knowledge that, e.g., tomatoes cause red_{exp} sensations.

Advocates of the knowledge argument grant that the informational assumption gives Mary knowledge of every proposition necessitated by \mathcal{P} . But the dualist and physicalist disagree about which propositions these are. The physicalist claims that experiential propositions are necessitated; the dualist denies. In order to wield the informational assumption against the knowledge argument, we must apply it only to propositions the dualist admits \mathcal{P} necessitates. I introduce a heuristic - *the zombie heuristic* - for determining which propositions these are. To determine whether the dualist grants that \mathcal{P} necessitates a proposition, check whether the proposition is true at both the actual and zombie worlds. The *zombie world* is an austerely physical duplicate of the actual world that contains no experiential properties. Zombies walk the walk and talk the talk (they even do zombie philosophy), but there is nothing "it is like" to be a zombie.

Our basic strategy is to play along with the dualist, using the zombie world to get him to agree that certain propositions are necessitated by \mathcal{P} , and thus knowable by Mary. We then use Mary's knowledge to refute the claim that Mary cannot

know that tomatoes cause red_{exp} sensations.²⁹ Consider some applications of the heuristic. Rocks exist at both the actual and zombie worlds; the dualist grants that \mathcal{P} necessitates that rocks exist. By the informational assumption, Mary knows that rocks exist. More interestingly, the dualist acknowledges that people in both the zombie and actual worlds enter neural state N whenever they see a ripe tomato. Thus \mathcal{P} necessitates the proposition that a person enters neural state N whenever they see a ripe tomato; Mary can know it. On the other hand, the dualist denies that, at both the actual and zombie worlds, people experience red_{exp} whenever they enter neural state N. The zombie heuristic tells us we cannot assume, via the informational assumption, that Mary knows this.³⁰

I offer a brief clarification of how I use the term 'dualist'. There are many varieties of dualist. Not all of them accept the metaphysical possibility of the zombie world. Nor are all dualists proponents of the knowledge argument. Here, when I speak of "the dualist", I intend those who, at the least, deny the operationalized modal definition of physicalism from page 2. The knowledge argument, as I construe it here, attacks this modal claim and thereby physicalism itself. My target here is the dualist who attempts to use the knowledge argument to yield anti-physicalist results about necessitation relations between the physical and the experiential. However, even this type of dualist need not accept the metaphysical possibility of zombies. However, they must accept the metaphysical possibility of worlds that are like the zombie world in the sense that they are austere phys-

²⁹Use of the zombie heuristic does not require that the zombie world be metaphysically possible.

³⁰The zombie heuristic is fallible. A proposition might be true at both the actual and zombie worlds, but false at some other world where \mathcal{P} is true. Maybe there are exotic worlds at which \mathcal{P} is true but rocks do not exist. This is unlikely, but a proposition's truth at the actual and zombie worlds does not guarantee its necessitation by \mathcal{P} . Here, we apply the informational assumption only to propositions that are plausibly necessitated by \mathcal{P} and granted by dualists as such.

ical, but not complete, duplicates of the actual world. Any dualist who uses the knowledge argument to establish such possibilities must accept the informational assumption. We can then use something like the zombie heuristic to wield the informational assumption against them. Rather than concoct a variety of heuristics for each variety of dualist, I'll lump them all together and pretend they all accept the possibility of zombies. What really matters is that all such dualists, in order to wield the knowledge argument, must grant the informational assumption. The zombie heuristic is merely a method for gauging what knowledge the informational assumption gives Mary.

The informational assumptions gives epistemological considerations metaphysical bite. It allows dualists to infer from a lack of epistemic implication (and Mary's inability to know) to a lack of metaphysical necessitation. But the informational assumption is a two-edged sword. It also cuts the other way. I'll use the informational assumption to infer from the fact that a truth is necessitated to the claim that the truth is implicated by the austere physical truths. I'll argue that, in fact, enough truths are implicated in this way to get Mary, while still in the black and white room, knowledge of all the truths about conscious experience.

3.9 Lonely Mary and the Pincer Argument

In order to draw metaphysical conclusions from what Mary can and cannot know, we must insist that Mary has conceptual mastery of experiential concepts, including RED_{exp} . In the knowledge argument as originally described, Mary did not have such mastery. But there is an obvious suggestion for how to proceed. If Mary did have conceptual mastery of RED_{exp} , would she be able to know the troublesome experiential propositions?

There are a variety of ways we could give Mary conceptual mastery of expe-

riential concepts. We might let her out of the black and white room and allow her to experience red_{exp} (and other colors), then re-capture her and give her an amnesia pill which eradicates her knowledge of what color experience tomatoes cause.³¹ She'd then have experienced red_{exp} , $\text{green}_{\text{exp}}$, and blue_{exp} , and plausible have mastery of the associated concepts. But she would not yet know which of those color experiences is caused by tomatoes, and which by the sky. If she can't come to know, despite her conceptual mastery, then the knowledge argument would survive my criticisms thus far. We might also give Mary conceptual mastery by performing neuro-surgery that allowed her to imagine color experiences.

Allowing Mary out of the black and white room, or performing neuro-surgery on her, introduces complications that I wish to avoid. The staunch defender of the knowledge argument might deny both the experiential and imaginative requirements on conceptual mastery of RED_{exp} . In order to give my argument widest possible appeal, I adopt the following strategy. I deprive myself of argumentative resources to avoid taking for granted anything with which the dedicated defender of the knowledge argument could take issue. In particular, I will not assume that conceptual mastery of a phenomenal concept requires either having had experiences of or the ability to imagine the relevant qualities.

Here is the most difficult Mary case I have been able to cook up. Suppose that Mary (a) is the only person ever to have existed, (b) never leaves the black and white room, (c) has never experienced a red_{exp} sensation, (d) cannot imagine a red_{exp} sensation, but has conceptual mastery of red_{exp} nonetheless.³² By making

³¹The use of "re-captured Mary", or a version of Mary who has had the relevant experiences, appears in Lewis (1988), Nida-Rumelin (1996), and ?. They all use re-captured Mary as a method of enabling Mary to possess phenomenal concepts. Here, I'm suggesting re-capturing Mary as a method for giving her conceptual mastery, not mere possession.

³²In Rabin (2011a) (sections 9-10), I discuss Mary cases in which Mary is not the only person

these assumptions I make it exceedingly difficult to argue that Mary can know. But I shall argue that even having granted the proponent of the knowledge argument this much, Mary can still know all the experiential propositions. All Mary needs is the recognitional ability to recognize a red_{exp} experience were she to have one, along with the knowledge granted her by the informational assumption. (Of course, there is a mystery here about how Mary came to acquire conceptual mastery (or even possession) of the concept RED_{exp} . Remember the dialectic: I wish to make no assumptions about how Mary acquires mastery and only minimal assumptions about what mastery requires, so that opponents cannot take issue with such assumptions.)

Before launching into the pincer argument, just below, I offer a comment for the reader. The pincer argument is difficult to understand. I'll present the material as clearly as I am able. However, the confused, annoyed, or impatient reader should feel free to skim this material and/or skip ahead to sections 10 and/or 11. In section 10 I consider objections to the pincer argument and the paper in general. In section 11 I expand the discussion and consider some big picture issues and the overall philosophical upshot.

The *pincer argument* goes as follows:

(1) Mary can know that tomatoes cause her to think a thought of the form THIS SENSATION IS R.

(2) Mary can know that the semantic value of the 'R' part of this thought is the

ever to have existed, Mary leaves the black and white room, and Mary experiences red_{exp} . As it turns out, in such cases it is surprisingly easy for Mary to know the relevant experiential truths by relying on the testimony of others, her expedition outside the black and white room, and her episode of experiencing red_{exp} , respectively. Readers interested in those cases should consult sections 9-10 of Rabin (2011a). The case that matters most is the one I am about to consider, in which assumptions (a)-(d) make it more difficult to argue that Mary can know.

concept RED_{exp} .

- (3) If Mary can know that tomatoes cause her to think THIS SENSATION IS R, and that 'R' means RED_{exp} , then she can know that tomatoes cause her to think THIS SENSATION IS RED_{exp} .
- (4) If Mary can know that tomatoes cause her to think THIS SENSATION IS RED_{exp} , then she can know that tomatoes cause red_{exp} sensations.³³
- (5) Therefore, Mary can know that tomatoes cause red_{exp} sensations.

I briefly sketch the support for each premise here; then I go into detail. The informational assumption ensures that Mary knows lot of facts about brain states and about how those brain states relate to representational states, especially to the syntax of those representational states. These facts will get Mary to the knowledge attributed to her in premise (1). Mary will integrate knowledge about her brain state at a time and knowledge about what she was thinking at that time to acquire the knowledge premise (2) attributes to her. Premise (3) is obvious. Mary can know that she is not in a deviant case, which is all that is needed to acquire the knowledge attributed to her in premise (4).

Since Mary has conceptual mastery of RED_{exp} , the recognitional requirement entails that Mary is disposed to think THIS SENSATION IS RED_{exp} (not THIS SENSATION IS $BLUE_{exp}$) whenever she sees a ripe tomato. If Mary can learn that seeing a tomato causes her to think THIS SENSATIONS IS RED_{exp} , then she can figure out that tomatoes cause red_{exp} sensations. By the recognitional requirement, concep-

³³The causation here must be of the appropriate type. For example, if tomatoes miraculously cause Mary to think THIS SENSATION IS RED_{exp} whenever they strike her forehead, there is no reason to think that tomatoes cause red_{exp} sensations. I leave the qualification about causation implicit.

tual masters of red_{exp} are disposed to think THIS SENSATIONS IS RED_{exp} if and only if they're having a red_{exp} sensation. Mary learns that tomatoes cause her to think THIS SENSATION IS RED_{exp} by combining two bits of information. First, she'll learn that tomatoes cause her to think THIS SENSATION IS R (premise (1)). Second, she'll learn that the semantic value of the 'R' portion of this thought is her concept RED_{exp} (premise (2)).³⁴ Premise (3) says Mary can combine these two pieces of information in the obvious way.

The proposition that tomatoes cause Mary to think THIS SENSATION IS R is necessitated by \mathcal{P} . By the informational assumption, Mary knows it. Evidence for this necessitation comes from considering the zombie world. Zombie Mary also thinks a thought of the form THIS SENSATION IS R. The difference between Mary and zombie Mary is that Mary's mental type 'R' has as its semantic value RED_{exp} , whereas zombie Mary's 'R' means $\text{RED}_{\text{zombie}}$. Actual speakers and zombies both mean THIS, IS, and SENSATION by their uses of 'this', 'is', and 'sensation', respectively. The dualist should grant that these semantic facts are necessitated by \mathcal{P} , and grant that \mathcal{P} necessitates that tomatoes cause Mary to think THIS SENSATION IS R.³⁵ By the informational assumption, Mary can know that tomatoes cause her to think THIS SENSATION IS R. Premise (1) is true.

To complete her epistemic route to knowledge that tomatoes cause Mary to think THIS SENSATION IS RED_{exp} , Mary needs only to learn that the 'R' component means RED_{exp} . The informational assumption ensures Mary knows the state

³⁴When I speak of 'semantic value' here, I mean (something like) Fregean sense, or meaning, not extension. On this terminology, the semantic values of 'this' and of 'Mary' differ, even on an occasion where a token of 'this' demonstrates Mary.

³⁵In section 10, I address how the argument fares if the dualist denies that some of these facts are necessitated. In particular, I consider the objection that the zombies' use of 'sensation' does not mean what we mean by 'sensation'.

of her own brain at every moment in history, both past and future. She also knows how her brain state relates to, and realizes, certain features of her representational or intentional state. For example, the syntactic structure of Mary's representational state is plausibly shared by zombie Mary, and necessitated by \mathcal{P} . Therefore, Mary knows the syntactic structure of her representational state at every moment in history, and she knows which aspects of her brain realize which aspects of that syntactic state. It follows that Mary knows exactly what brain state-type realizes the 'R' token-type of her representational state. Mary knows exactly at which moments she has thought, and will think, an 'R' token, because she knows when her brain is in the state that realizes the 'R' token.

Mary needs to figure out the semantic value of her 'R' tokens. 'R' might mean RED_{exp} , or $BLUE_{exp}$, or $MANGO$, or it might mean something else. Mary figures out which of these concepts 'R' means by thinking each concept in turn. By the informational assumption, Mary already knows the neural correlate of 'R' tokens. She can tell, on the basis of \mathcal{P} alone, when she's actively considering a proposition containing 'R'. (I'm assuming a neurologically realized difference between actively considering a proposition and merely believing it). When Mary actively considers the proposition $THIS\ MANGO\ IS\ A\ SUCCULENT\ MANGO$, she does not actively consider a proposition containing an 'R' component. She'll deduce that 'R' does not mean $MANGO$. When Mary considers the proposition $ALL\ RED_{exp}\ THINGS\ ARE\ RED_{exp}$ (almost any thought containing RED_{exp} will do), she does actively consider a proposition containing 'R'. She'll deduce that 'R' means RED_{exp} . Therefore, premise (2) is true. (This argument requires a very minimal and tremendously plausible assumption about knowing one's own thoughts. It requires the assumption that when Mary considers the proposition $ALL\ RED_{exp}$

THINGS ARE RED_{exp}, she can know that she is considering a proposition containing RED_{exp}, and not MANGO or BLUE_{exp}.)

Premise (3) is obvious. It says that if Mary can know both that tomatoes cause her to think THIS SENSATION IS R and that 'R' means RED_{exp}, then she can know that tomatoes cause her to think THIS SENSATION IS RED_{exp}.

Premise (4) says that Mary can infer from the fact that tomatoes cause her to think THIS SENSATION IS RED_{exp} to the fact that tomatoes actually cause red_{exp} sensations. To do this, Mary needs to check that she is not being misled in the particular case of tomatoes. For example, if Mary applied the concept RED_{exp}, for the most part, to all and only red_{exp} experiences, but misapplied to tomato-caused blue_{exp} sensations (perhaps because tomatoes trigger her allergies), then tomatoes would cause Mary to think THIS SENSATION IS RED_{exp} without tomatoes actually causing red_{exp} sensations. Mary knows she is not in such a deviant case. She knows that tomatoes, strawberries, and fire engines all emit light of wavelength 700 nanometers, have dramatically similar effects on the color processing component of her visual system, and generate the same sensational neural correlate. She knows tomatoes do not trigger an allergic reaction that causes her system to misfire and label her experiences in a non-standard way. Premise (4) is true.

Premises (1)-(4) of the pincer argument are true. The argument is valid. Therefore the conclusion is true: Mary can know that tomatoes cause red_{exp} sensations. By combining (i) knowledge "from below" about brain states and the syntax of the representational states those brain states realize (knowledge allotted to Mary by the informational assumption) and (ii) knowledge "from above" about what thought (ALL RED_{exp} THINGS ARE RED_{exp}) she was thinking at a given time, Mary

can know that tomatoes cause her to think THIS SENSATION IS RED_{exp}.³⁶ From there, it's only a short step to knowledge that tomatoes cause red_{exp} sensations.

3.10 Objections

In this section, I consider five objections. The first two are objections to claims about intentionality used in the pincer argument; the third is an objection to the recognitional requirement on conceptual mastery of RED_{exp}; the fourth and fifth are objections to my presentation of the knowledge argument.

One objection to the pincer argument says that the intentional facts I claim Mary knows are not necessitated by \mathcal{P} , and thus the informational assumption does not give Mary knowledge of them – premise (1) is false. The important question here is “Are the intentional propositions needed by Mary to learn that tomatoes cause red_{exp} sensations necessitated by \mathcal{P} ?” If the answer is ‘yes’, then these propositions are, by the informational assumption, known by Mary, and the objection is answered. The objector, a dualist, claims the answer is ‘no’.

One view that supports this objection claims that all intentional facts depend on facts about conscious experience – phenomenology is prior to intentionality. Something like this view is suggested in Horgan & Tienson (n.d.), Graham *et al.* (2007), and Strawson (1996). If this is correct, and zombies are possible, then there is no intentionality at the zombie world, no representational states are necessitated by \mathcal{P} , and no intentional facts are knowable by Mary via the informational assumption. However, the “phenomenology first” view couples badly with dualism. If you believe that zombies are possible, you should believe that at least

³⁶The “pincer argument” gets its name from Mary’s two-pronged approach to knowledge that tomatoes cause her to think THIS SENSATION IS RED_{exp}. In military strategy, a simultaneous attack from two sides is called a ‘pincer movement’ or ‘double envelopment’.

some intentional facts are independent of phenomenology. Zombies appear to utter sentences and think thoughts. Intentional explanations and idioms will be tremendously successful in predicting and explaining the behavior of zombies. This does not guarantee that zombies have intentionality or represent the world, but it is very good evidence. (Similarly, the great success of the heliocentric model of the universe in predicting and explaining the motion of the planets does not guarantee its truth, but it provides very strong evidence). At least some zombie sentences and thoughts are true, others false. If so, then at least some intentionality floats free of consciousness. For example, it is very plausible that the syntactic form of sentences, and of thoughts, does not depend on conscious experience, and thus does not vary between Mary and zombie Mary. Furthermore, it is plausible that the meaning of many terms, such as 'is', 'that', and 'basketball', does not shift between the actual and zombie worlds. If these two claims are correct, then the pincer argument can fend off the objection. That argument required only that \mathcal{P} necessitates the proposition that tomatoes cause Mary to think THIS SENSATION IS R. The dualist should admit that enough intentionality is independent of phenomenology to make this proposition true at the zombie world. If so, the proposition is plausibly necessitated by \mathcal{P} and knowable by Mary, as premise (1) claims.

The second objection, closely related to the first, claims that the meaning of 'sensation' varies between the actual and zombie worlds. Since zombies have no phenomenal experiences, they mean something different by 'sensation'. But I claimed that 'sensation' has the same meaning in our mouths that it does in zombie mouths. I reply that the claim that 'sensation' has the same meaning at the actual and zombie worlds is inessential to the pincer argument. Even if the

meaning shifts, Mary can learn the meaning of 'sensation' using the same trick she used to learn the meaning of 'R'. By the informational assumption, she'll know that tomatoes cause her to think THIS S IS R. She can then use the method previously described (of actively considering thoughts containing the concept RED_{exp}, or SENSATION), to deduce that 'S' means SENSATION.

This reply offers a lesson and bolsters the reply to the first objection. The pincer argument requires only that facts about the syntactic structure of Mary's representational states are necessitated by \mathcal{P} (and thus knowable by Mary according to the informational assumption). Once Mary knows she is thinking a thought with a given syntactic form, and is able to re-identify mental language tokens as being of the same type, she can use the familiar trick to determine the semantic values of those tokens. By "the familiar trick", I mean the process by which Mary deduces the semantic value of some syntactically individuated neural state. Mary knows that some (potentially vastly complicated) neural state N has a semantic value. She does not know which value, but she can deduce the value by checking - by "the familiar trick". Does N mean MANGO? To find out Mary need only actively consider a MANGO thought and check whether, when she did so, her brain entered the N state. If it did not, then N does not mean MANGO. This is exactly the trick Mary used to determine that 'R' meant RED_{exp}. This syntactical information about brain states is knowable by Mary because it is tremendously plausible that the syntax of an agent's language (both spoken and mental) is necessitated by \mathcal{P} and shared by Mary and her zombie twin.

The pincer argument relies on the informational assumption and Mary's conceptual mastery of RED_{exp}. Only the recognitional requirement on conceptual mastery - not the experiential or imaginative requirement - is needed. Using

only the recognitional requirement gives the argument wider appeal. Deniers of the experiential and imaginative requirements can be convinced. But the true skeptics will remain doubtful. Why think that conceptual mastery of the experiential concept RED_{exp} requires so much? Why think that recognitional capacities are a requirement on conceptual mastery of RED_{exp} , or on phenomenal concepts in general? I do not have a knock-down argument against the dedicated skeptic. I ask the reader to consider a case, and consult his or her intuition.

Conceptual mastery is the type of understanding had by the proverbial “experts” to whom users who merely possess the concept, but do not have mastery, defer. The thought contents of mere possessors are determined by those who have full-fledged conceptual mastery. Imagine a case in which conceptual mastery of RED_{exp} does not require the ability to recognize a red_{exp} experience as RED_{exp} . Suppose we are all blind, but use the concept RED_{exp} , and defer in our use to the sighted master, whom we call ‘Sensei’. The reference, content, and Fregean sense of our RED_{exp} concept is determined by Sensei’s. Sensei, when presented with three color sensations, one red_{exp} , one $blue_{exp}$, and one $green_{exp}$, and asked, “which of these sensations is red_{exp} ?” will throw up his hands and say, “I have no idea!” Given that Sensei is the ultimate authority here (he does not defer to further experts), it’s unclear what makes the concept RED_{exp} have red_{exp} rather than $blue_{exp}$ sensations in its extension. One way to ensure that RED_{exp} refers to red_{exp} sensations is to have the expert to whom everyone defers identify red_{exp} sensations as falling under RED_{exp} , and $blue_{exp}$ sensations as not. Conceptual mastery is that which is had by the proverbial experts. The experts regarding RED_{exp} must be able to recognize red_{exp} sensations as such; otherwise they aren’t experts, and don’t have conceptual mastery.

An objector might disagree with my interpretation of the knowledge argument. He or she might think that I've targeted the wrong propositions and claim that there is some other proposition that Mary does not know. I disagree: there is no such proposition. By the informational assumption, Mary has a tremendous amount of information at her disposal. This information will allow her to know any proposition you like, often by means similar to those I have outlined.

An objector might instead maintain that the knowledge Mary lacks is not propositional. I have no beef with this claim. I say only that there is no propositional knowledge that Mary lacks.³⁷ If the knowledge Mary lacks is not propositional, it is difficult to see how her ignorance entails the falsity of the physicalist's necessitation claim. Necessitation is a relation between propositions (or perhaps between states of affairs, which are easily translated into the language of propositions). The objector would need to explain how Mary's lack of non-propositional knowledge entails the falsity of physicalism construed as a claim about necessitation.

3.11 Taking Stock

One upshot here is that Mary, sitting in her black and white room, bestowed with conceptual mastery of experiential concepts such as RED_{exp}, BLUE_{exp}, and TICKLISH_{exp}, *can* know every true proposition involving those concepts. This is a tremendously surprising result. It is worth noting that the result does not depend on the claim that physicalism is true. Even if dualism is true, Mary will *still* be able to know the experiential propositions.

³⁷More accurately, I claim that there is no propositional knowledge Mary lacks in a way that challenges physicalism. There is no knowledge that Mary lacks for reasons other than lack of concept possession and/or conceptual mastery.

My position is very different than traditional phenomenal concept strategies, and I do not think my view is deserving of the name or the lineage. All phenomenal concept strategies accept that (a) Mary cannot know certain experiential truths on the basis of the physical truths and (b) the physical truths necessitate these experiential truths nonetheless. Special aspects of “phenomenal concepts” are then wheeled in to explain how (a) and (b) could both obtain. Phenomenal concepts explain how there could be necessitation from the physical to the experiential without epistemic implication. My strategy is very different. I claim that once we give Mary the conceptual capacities necessary to draw any metaphysical conclusions, Mary *will* be in a position to know the experiential truths. Thus I deny the datum (a) with which phenomenal concept theorists begin.

The most popular response to the knowledge argument is to deny the modal rationalist claim that one can draw metaphysical conclusions about necessitation from epistemological considerations. Phenomenal concepts supplement this response. Special features of phenomenal concepts are wheeled in to explain how there can be “opaque” necessitation – necessitation without epistemic implication.

My response is much different. I’m ready to spot the advocate of the knowledge argument his modal rationalism. Let’s assume that one *can* draw metaphysical conclusions on the basis of some epistemological notion.³⁸ Even then, I argue, we can preserve our physicalism in the face of the knowledge argument. This gives my account two dialectical advantage over most extant responses to

³⁸Of course, I’m not willing to spot the advocate of the knowledge argument any inference he likes. Section 6 demonstrates that, if the modal rationalist claim that metaphysical conclusions follow from epistemology is to have any chance of being true, we must insist that the epistemic agents have conceptual mastery. The rationalist position I’m willing to play along with must acknowledge at least this much.

the knowledge argument.³⁹

First, a modal rationalist can endorse my reply, and not most others. Second, since most responses deny the modal rationalist connection between metaphysics and epistemology, the debate quickly shifts into deep issues in the philosophy of modality. With this shift comes the risk of devolution into a clash of basic intuitions about links between modality, knowability, a priority, and conceptual truth. I avoid the devolution by granting the modal rationalist his connections between epistemology and metaphysics, while still offering the physicalist an adequate reply to the knowledge argument.

Finally, I take pain not to overstate my position. Some philosophers, most notably Chalmers (1996b) and Jackson (1998), interpret the knowledge argument as an attempted demonstration that there is no *a priori* implication from \mathcal{P} to the experiential truths. Have I demonstrated that there is such an *a priori* implication - at least for a subject who has conceptual mastery? This question is difficult. The answer depends on the relationship between *a priori* knowledge and introspection, and whether certain types of introspection can yield *a priori* justification.

In the pincer argument, Mary used a four step procedure to figure out the RED_{exp} facts. First, by the informational assumption, Mary knew that tomatoes, fire engines, and the like cause her to token 'R' representations. Second, she needed to determine that Rs are RED_{exp} representations and not BLUE_{exp} representations. She did this by actively thinking a thought involving the RED_{exp} concept. At the moment she does so her brain realizes a R token. From this she learns that R tokens are tokens of RED_{exp} (not of BLUE_{exp}). Importantly, Mary had to actively consider a thought involving RED_{exp} at time t and use her knowledge that

³⁹Thanks to Ned Block for encouraging me to advertise this selling point and stress the differences between my account and those already in the literature.

she considered that thought (a thought containing RED_{exp}), and not another, at t . This move is certainly introspective. The relevant question is whether the type of introspection by which one determines what thought one is thinking, and what concepts that thought contains, yields a priori, or merely a posteriori, justification.

I cannot address here the difficult questions surrounding the relationship between a priori justification and introspection. One move available to Chalmers, Jackson, and others, is to *define* a notion of a priori justification that does not allow introspection of the type at issue here. They can then use Mary as evidence that there is no a priori_{no introspection} implication from the austere physical facts \mathcal{P} to the experiential facts, and attempt to draw metaphysical conclusions. I admit that I have not shown that there is an a priori_{no introspection} route from \mathcal{P} to the experiential facts, even if the agent has conceptual mastery. However, this stipulative move does not resolve the debate. The knowledge argument requires the existence of a link between epistemic implication and metaphysical necessitation. The a priori version of the knowledge argument requires an important (potentially constitutive) link between a priori implication and metaphysical necessitation. Defenders of the a priori knowledge argument can stipulate whatever sense of “a priori” they like. But they are not free to stipulate that it is a priori_{no introspection}, rather than some *other* conception of the a priori, that has the vital link to metaphysical necessity. This latter claim requires argument.

For now, I leave it open (i) whether there is any notion of a priori implication that allows the drawing of conclusions about necessitation and (ii) if there is such a notion, whether it does or does not allow introspection. If it does not, then a strict a priori version of the knowledge argument might survive my criticisms. In

this case, I will have shown only that there is an *armchair* implication from \mathcal{P} to the experiential facts. This result is itself significant. If the relevant notion of the a priori does allow introspection of the relevant type, then I take my arguments to demonstrate that there is an a priori implication from the complete physical facts \mathcal{P} to the experiential facts, and the knowledge argument is sunk.

3.12 Conclusion

The knowledge argument has a tough row to hoe. It requires a perfect storm of epistemological and metaphysical conditions. It requires that Mary have the experiences and capacities necessary to possess and master phenomenal concepts. And the argument requires that Mary simultaneously be unable to know experiential propositions and, by the informational assumption, able to know every proposition necessitated by the complete austere physical truth \mathcal{P} . That's a lot to require.

The informational assumption gives Mary a lot to work with - information about the neural correlates of sensations, about what every human being has said throughout the history of time, about the relationship between brain states and intentional or representational states, and a tremendous deal more to boot. With such a plethora of information, it is not difficult to figure out the experiential truths themselves. Even in variations on the knowledge argument in which no one speaks the word 'red', or in which Mary is the only person ever to have existed, Mary can use the knowledge allotted to her by the informational assumption to figure out the experiential truths.

We could continue doing variations on the argument, on Mary's situation, and on what proposition cannot be known, until the cows come home, but these basic constraints remain. We cannot draw any metaphysical conclusions unless

Mary has conceptual mastery of experiential concepts. But once she has conceptual mastery, Mary will be able to know all the experiential propositions she likes. She'll know about the redness of a ripe tomato, the greenness of a grassy field, and the brilliant blue of a clear sky on a cold winter day. The knowledge argument is a beautiful and provocative piece of philosophy. Jackson sketches a very promising line of anti-physicalist attack. But when it comes to the details, it is surprisingly difficult to fill in the sketch, and make good on Mary's metaphysical potential.

Chapter 4

Toward a Theory of Conceptual Mastery

Chapter Abstract

This paper investigates the question “Under what conditions does a thinker fully understand, or have mastery of, a concept?” A thinker can possess, and think thoughts with, concepts he or she does not fully understand. I argue against three views of conceptual mastery, according to which conceptual mastery is a matter of holding certain beliefs, being disposed to make certain inferences, or having certain intuitions. None of these attitudes is either necessary or sufficient for conceptual mastery. I propose and respond to objections to my own “meaning postulate view” of the conditions under which a thinker has mastery of a concept.

4.1 Introduction

Concepts are the basic units of thought. The thought OSTRICHES LIKE CHOCOLATE is composed of the concepts OSTRICHES, LIKE, and CHOCOLATE. An agent *possesses* a concept when he or she can think thoughts of which the concept is a component. An agent has *conceptual mastery* of a concept when he or she fully understands that concept. One can possess a concept without having mastery. Many people use technical concepts that have worked their way into public consciousness without fully understanding them (examples include DARK MATTER, CHAOS THEORY, CLOUD COMPUTING, TRANSISTOR, ID, FRACKING, and HEDGE FUND). Any theory of conceptual mastery must answer the following fundamental question: “Under what conditions does an agent have conceptual mastery?” This is the question I hope to answer, at least in part, in this paper.

I will argue for “the meaning postulate view” of conceptual mastery, according to which mastery of a concept is a matter of taking certain core rules to govern the use of the concept. These rules, are, in effect, “meaning postulates” that help give the concept its meaning.¹ I will argue against “the belief view”, “the inference view”, and “the intuition view”, according to which mastery of a concept involves believing, inferring, or intuiting in accordance with certain propositions or inferences.

I now offer a roadmap of the paper. In section 4.2 (“Preliminaries”) I clarify terminology, lay out assumptions, argue briefly against skepticism about conceptual mastery, and canvas four potential views about what it is to master a concept. Each of the three views against which I argue maintains that some set of attitudes

¹I use ‘meaning’ in the generous sense, to cover any variety of representational quality, not just the representational features of language. On this approach, non-linguistic mental states, such as (some) conceptual thoughts and perceptual states, have meaning.

(beliefs, inferences, or intuitions) is necessary and sufficient for mastery of a concept. In sections 4.3 (“Deviant Masters”) and 4.4 (“The Special Status Problem”) I argue against the necessity and sufficiency (respectively) of these attitudes for conceptual mastery. The failures of these views point the way toward a successful theory, which I present in section 4.5 (“The Meaning Postulate View”). Section 4.6 (“Three Objections”) considers and replies to objections to the meaning postulate view.

4.2 Preliminaries

4.2.1 Terminology and assumptions

In this section, I’ll lay out assumptions, clarify terminology, and sketch a framework in which to commence inquiry. First, I use the term ‘proposition’ to describe a structured composite of concepts. For example, the proposition OSTRICHES LIKE CHOCOLATE is composed of the concepts OSTRICHES, LIKE, and CHOCOLATE. I use small capitals, e.g. OSTRICH, to express elements at the conceptual level, either concepts or complex propositions. Following Frege (1952/1892b), I will also use the term ‘thought’ to denote propositions so described. I adopt a broadly Fregean approach to concepts and propositions. This approach is not intended to be controversial. Inevitably, due to prior commitments about the nature of propositions, some readers will take it as such. Readers who bridle at my use of ‘proposition’ should substitute the word ‘propersition’ for every occurrence of ‘proposition’. Propersitions are the complete thoughts I am currently describing. I assume that one can, and for many purposes should, type mental states using propositions (propersitions) so described.

A certain class of mental states (the so-called “propositional attitudes”) can

usefully be categorized by a combination of a *mode* and a proposition. The mode characterizes (at least one of) the thinker's attitude(s) toward the proposition. Examples of modes include belief, desire, and fear. Examples of complete propositional attitudes include a belief that ostriches like chocolate, a desire that ostriches like chocolate, and a fear that ostriches like chocolate. I don't deny that there are lots of other ways to type and describe mental states and propositions, or that those ways are useful, informative, or illuminating. I assume only that the locutions I've described constitute one useful way of thinking and talking about contentful mental states. Something like the above approach should be compatible with and fruitful for a wide variety of philosophical and psychological projects. I'll assume this approach throughout the paper.

I assume that there is a distinction between merely possessing a concept (i.e. being able to take some attitude toward a proposition that contains the concept) and fully understanding that concept.² Not uncontroversially, I assume also that, sometimes, thinkers entertain thoughts containing concepts they do not fully understand, and thus possess those concepts without mastery of them. Theorists of content and of concepts tend to assume, often implicitly, without argument, and without recognizing that any such assumption has been made, that possession of a concept (i.e. the ability to think thoughts containing the concept) entails mastery of that concept. Often they fail to note any distinction at all between possession and mastery. The difficulty is partly due to ambiguity in the terminology of "having/possessing a concept". That locution can be used to express either a relation of possession or a relation of mastery between a thinker and a concept.

²For further discussion of this distinction, cf. Burge (1979, 1986), Greenberg (MS): 13-4, *passim*, Greenberg (2001): 28-9, *passim*, Greenberg (2005): 300-1, 307, Greenberg (2014), Bealer (2002a): 221-222.

Here, I reserve the locution “x possesses the concept C” to mean that x can think thoughts that contain C. x may or not may have mastery of C.

In so far as theorists do recognize the possibility of concept possession without mastery, they often chalk up the phenomenon to and add a proviso about “deference”, and then continue to ignore the possibility of possession without mastery. Here, I do not wish to tackle issues about the relation between concept possession and conceptual mastery and their relations to the theory of content and of concepts more generally. (I point the reader to the discussions of Greenberg (MS, 2001, 2014), Burge (1979, 1986), and Peacocke (1992, 1999, 2008)). Instead, I will simply assume that thinkers can possess concepts that they have not mastered. Given that assumption, I then seek to answer the question, “What does it take for a thinker to master, or fully understand, his or her concept?”

I take no official position on what is required to possess a concept. I require only that whatever is required to possess a concept, it does not always entail that one have mastered the concept. However, I believe (but will not assume) that the requirements on concept possession are relatively minimal. It’s easy to acquire a concept, especially when that concept is expressed by a public language term. Hearing someone utter a term that expresses a concept not previously possessed, and then using that term and concept in thought, without intending to appropriate the term and give it new meaning, will often be sufficient for acquisition of the concept. For example, if I hear a doctor declare that the patient has lupus, and begin believing that the patient has lupus, then I can come to possess, but not master, the concept LUPUS. This is so despite my ignorance regarding LUPUS. I know little about LUPUS other than that it refers to some type of ailment.

Most controversially, I assume a distinction between “core” elements associ-

ated with a concept and “non-core” elements. These elements might be propositions, inferences, mental transitions, rules, or something else. Core elements are in some sense especially important to the meaning of the concept. They play a special role in the concept’s having the representational qualities that it does and being the concept that it is. One might further claim that core elements partly or wholly constitute the nature of the concept. For example, I take the proposition ALL BACHELORS ARE MALE to be a “core” proposition for the concept BACHELOR. ACHILLES IS A BACHELOR is non-core. One might also maintain that core propositions are “analytic” or contribute to a priori justification. I remain agnostic on this count. To accept the distinction between the core and the non-core, one must resist, at least to some degree, the arguments of Quine (1951) against the analytic/synthetic distinction.

4.2.2 Cores and Conceptual Roles

According to a broad philosophical tradition called the “conceptual role approach” to mental content and concepts, a concept’s role in thought, and especially its connections to other concepts, plays a vital role with respect to the concept. For example, according to Greenberg (MS)’s characterization, conceptual role theories (which he calls ‘inferential-role theories’) maintain that “what makes it the case that a thought involves a particular concept is that the thinker deploys a disposition to move between mental states that is appropriately isomorphic to the concept’s canonical or defining connections to other concepts.” (21) The main rival to conceptual role theories are informational/covariational accounts, which maintain that a concept’s reference, rather than its relations to other concepts, plays the vital role.

My use of the phrase ‘a/the vital role’ is intentionally vague, because extant

theories tend to run together a variety of theoretical roles that conceptual or inferential roles, or transitions between mental states more broadly, might play. I envision at least six. First, one might claim that conceptual roles partly or wholly individuate concepts. Second, one might claim that conceptual roles partly or wholly constitute the nature of the concept. Third, one might claim that a concept's role determines the extension of a concept. For example, a theory might say that the extension of the concept is that which makes the concept's canonical transitions truth-preserving. Fourth, one might tie concept possession to conceptual role, in the sense that possession of a concept requires having a mental item that plays the appropriate role. Fifth, one might further claim that conceptual roles make it the case that a particular mental state or thought contains one concept rather than another. (These last two are closely connected, but there is a difference. The former concerns a general ability (possession), while the latter pertains to the content of a particular mental state.) Sixth, one might claim that conceptual roles determine mastery of the concept, in the sense that one has mastery of the concept if and only if one appropriately grasps the concept's conceptual role or canonical transitions.

My project and sympathies fall with the conceptual role tradition, broadly conceived. My acceptance of a distinction between "core" and "non-core" elements, which I described as "propositions, inferences, mental transitions, rules, or something else", places me in this camp. The core elements are something like the "canonical transitions" on which conceptual role theories rely. All the accounts of conceptual mastery that I will discuss fall broadly within the conceptual role approach. (For an approach to conceptual mastery that bridges the conceptual role vs. co-variational/informational divide, cf. Greenberg (2014).)

It's worth laying my views on the table, partly to help the reader understand the angle from which I approach this inquiry. I believe that canonical transitions partly, but not wholly, individuate concepts (reference does some individuation as well). I believe that canonical partly constitute the nature of the concept. I do not believe that canonical transitions determine extension. I tie neither concept possession nor what makes a thought contain one concept rather than another, to conceptual role. My favored account of conceptual mastery, the "meaning postulate view" (section 4.5), says (roughly) that mastery of a concept amounts to understanding of the canonical transitions. I will not assume any of these beliefs in the discussions to follow. I will argue for (not assume) the claim that conceptual mastery involves understanding of canonical transitions. The three elements of conceptual role approaches that I endorse (concept individuation, nature and mastery) fit together. From the claim that concepts are partly or wholly individuated in terms of canonical transitions, it's a short step to the claim that those transitions form part of the concept's nature. It's then natural to claim that understanding the concept is a matter of understanding that nature, which entails understanding those canonical transitions.³

4.2.3 Precising the Notion of Conceptual Mastery

In this section, I clarify the target notion: mastery, or full understanding, of a concept. A thinker has mastery of the concept FLYING BUTTRESS exactly when she fully understands FLYING BUTTRESS. I believe that the notion of full understand-

³Greenberg (2014)'s account of conceptual mastery manifests a similar connection between nature and understanding of that nature. He writes (1,7) that mastery of a concept is a matter of "having a practically available understanding of its nature or essence."

ing, or having mastery, of a concept, is relatively pre-theoretical and everyday.⁴ We express the relation using locutions like ‘she fully understands the concept flying buttress’, ‘she grasps the meaning of the concept flying buttress’, and ‘she fully understands the meaning of ‘flying buttress’’.⁵ Mastering a concept is not an impossible feat, or an idealized state achievable only by hypothetical thinkers. You and I have mastery of many concepts.

Mastering a concept is not merely a matter of knowing many truths involving the concept. One might know many truths containing CHAIR, but if one fails to realize that chairs are for sitting, then one fails to master CHAIR. Mastery is tied to grasping certain important or key insights about the concept. Mastery of a concept is compatible with ignorance of important truths about that to which the concept refers. For example, I believe that mastery of the concept WATER is compatible with ignorance of the fact that water is H₂O.⁶ Someone who knows that water is H₂O knows more about water than someone who does not. She

⁴Greenberg (2014): 6-7, Greenberg (MS): 14, Greenberg (2001): 28-30 agrees that the notion of conceptual mastery, or full understanding, is pre-theoretic and everyday, roughly mirroring the notion of fully understanding or mastering a word.

⁵We must be careful here. It’s possible that fully understanding the linguistic meaning of a word and fully understanding the concept it expresses could come apart. Burge (1990) claims that this is possible, and attributes Gottlob Frege a similar view. The example Burge offers involves the mathematical concept LIMIT from the calculus. In the early days of the development of the calculus, both Isaac Newton and Gottfried Leibniz used the notion of a limit. Both possessed the concept LIMIT, but neither thinker had a clear grasp on it. The linguistic meaning of the word ‘limit’, however, is determined by how individual speakers have used the term. With respect to the linguistic usage of ‘limit’, Newton and Leibniz were the experts. They fully understood the linguistic meaning of ‘limit’. Yet they failed to fully understand the concept LIMIT that the word expressed. If this description of the case is accurate, then it seems possible to fully understand a term’s linguistic meaning yet fail to fully understand the associated concept. We might take the expression ‘x fully understands the meaning of term t’ to have two uses. It can be used to say that x fully grasps the linguistic meaning of t. Or it can be used to say that x fully grasps the concept that t expresses. Only when ‘x fully understands the meaning of t’ is used in the first sense does mastery of a concept come apart from full understanding of the meaning of a linguistic item. Throughout this paper, I’ll use the expression ‘fully understanding the meaning of term t’ exclusively in the second sense.

⁶Greenberg (2001): 134, 137-8 agrees.

also better understands water (the stuff) and its nature. But I deny that she better understands the concept WATER. That water is H₂O is a deep, necessary, and constitutive truth about water. But H₂O-ness is not part of the concept WATER itself.

Consider another case.⁷ Suppose that it turns out that all chairs are, in fact, artifacts left by an alien civilization. By an odd quirk of physics, the alien technology is necessary to create any object that can be sat on like a chair. Nothing can be a chair unless it is an alien artifact. If this is so, then “all chairs are alien artifacts” will be a deep and important truth about the nature of chairs. Nevertheless, I claim that one can fully understand and master the concept CHAIR without knowing about the chair-alien artifact link. Such masters of CHAIR will fail to fully understand chairs and their nature. But their understanding of the concept CHAIR need not be lacking. I take these claims about WATER and CHAIR to be intuitive. But even if one finds them controversial, they should help the reader grasp the notion of conceptual mastery as I conceive it.

4.2.4 Against Skepticism

In developing a theory of conceptual mastery, I take as my starting point the seminal discussions of Burge (1979, 1982, 1986). Burge deserves much credit for bringing into the philosophical arena the phenomenon of *incomplete understanding*. Consider Burge (1979)’s example of Alf. Alf approaches his doctor, complaining of arthritis in his thigh. Alf’s doctor corrects him, pointing out that one cannot have arthritis in the thigh. Arthritis, by definition, affects only joints. Alf accepts the correction, admits that he was mistaken to believe that he had arthritis

⁷This case is based on a thought experiment in Burge (1986): 263-64. I have altered the case and put it to a different use.

in his thigh, and modifies his usage of 'arthritis' and ARTHRITIS accordingly. The important point here is that Alf was able to think the thought I HAVE ARTHRITIS IN MY THIGH, and thereby possessed the concept ARTHRITIS, while significantly misunderstanding the concept. Alf incompletely understood ARTHRITIS. Examples of this nature can be cooked up for almost any concept whatsoever.

If agents can possess concepts while incompletely understanding them, it's natural to ask what agents must do to possess a concept and completely understand it - i.e. to have conceptual mastery. It's natural to think that such complete understanding is attainable by everyday people. Not everyone is like Alf.

Some will be skeptical that there is any such thing as "mastery" or "full understanding", or that such notions are theoretically significant or useful.⁸ I won't tackle such skepticism head on, but I want to make a few remarks to clarify what full-blooded skepticism about conceptual mastery must be and to lay out some considerations against it.

One can understand better or worse. Understanding is, in general, a matter of degree. Thus mastery of a concept will be a matter of degree. We might rank hypothetical understanders of a concept according to their understanding. On one end of the spectrum we have agents who meet only the minimal requirements (whatever they may be) for possession of the concept. On the other end of the spectrum we have agents who are omniscient with respect to the concept. They know every true proposition containing the concept and know all the important truths about the thing in the world to which the concept refers.

The strongest form of skepticism about conceptual mastery maintains that there are no theoretically significant breaks, marks, or groupings among the hy-

⁸Among the skeptics I count Ball (forthcoming) and Williamson (2008, draft), [p.c.].

pothetical agents so ranked. Instead there's a continuous spectrum. A weaker form of skepticism admits that there are theoretically significant groupings, but none of those groupings corresponds to the notion of conceptual mastery as I describe it. One initially tempting form of skepticism about conceptual mastery goes something like this. Every concept occurs in a wide variety of truths. Some of these truths are known by almost everyone who uses the concept. They're common knowledge. We use the term 'fully [completely] understands the concept C' as a label for those individuals who know all the common knowledge truths involving C. On this view, there is such a thing as conceptual mastery, but it marks only a linguistic accident.

I deny all three forms of skepticism. I believe that there are some theoretically significant breaks, marks, or groupings among the hypothetical agents described above. I believe that one (perhaps rough and ready) way to informatively group the agents is to divide them into those that have mastered the concept and those that have not.

I offer a brief argument against the skeptic. It relies on two claims. First claim: Incomplete understanding occurs. Second claim: With respect to a given concept, there is a distinction between core truths and non-core truths. "All bachelors are male" is core vis-a-vis BACHELOR. "No bachelors live on Mars" is not. Between two agents who are otherwise the same, there's a big difference, with respect to understanding of BACHELOR, between (i) the agent who fails to accept that all bachelors are male while accepting that no bachelors live on Mars and (ii) and the agent who fails to accept that no bachelors live on Mars but accepts that all bachelors are male. The second understands BACHELOR much better than the first. Once we have a theoretically significant distinction between core and

non-core truths, we can define up a theoretically significant level of conceptual mastery according to which mastery involves having a “grip” on the core truths. (We might also use core inferences, rules, or what have you). This tells against both the strong and weak forms of skepticism above.⁹ The linguistic form of skepticism I described is defeated by the above argument in combination with the observation that both “all bachelors are male” and “no bachelors live on Mars” are common knowledge.

One might resist the above arguments by maintaining that most concepts are not like BACHELOR. According to this line, BACHELOR permits a distinction between core and non-core elements, but most, and perhaps almost all, concepts do not. I admit that BACHELOR is special in certain respects, and in some ways not representative of concepts more generally. BACHELOR has a particularly “strong” core. One can reasonably maintain that the core elements determine the reference of BACHELOR, provide necessary and sufficient conditions for bachelor-hood, or give the meaning of BACHELOR. For many (perhaps most) concepts, analogous claims will not be as plausible.

But, importantly, many concepts other than BACHELOR yield core elements. Consider some examples. CHAIR: chairs are for sitting. ARTWORK: artworks are created. MUG: mugs function to hold liquid. WINE: wine is made from grapes. BLANKET: blankets have a large surface area to thickness ratio. CAVE: caves are enclosed spaces. SWORD: a sword has a blade. The method of examples cannot possibly show that even a significant minority of all concepts have cores. But hopefully the examples alleviate some of the fear that BACHELOR is idiosyncratic in having any core at all, and lend some support to the idea that a significant

⁹Greenberg (2001): 134, Greenberg (MS): 14-17, Greenberg (2014): 7-11 implements a similar strategy in arguing against skepticism about conceptual mastery.

portion of concepts have a core. The reader can further satisfy him or her self by doing a few cases on his or her own. My discussion of conceptual mastery and full understanding will not apply to those concepts which do not admit of a core vs. non-core distinction. (I do not assume that there are any such concepts). From here, I will speak as if all concepts have a core vs. non-core distinction, but one can take my comments as restricted to those concepts which do.

In sum, if one is (a) externalist enough to believe that there is such a thing as possessing and using a concept with incomplete understanding, and (b) resistant enough to the arguments of Quine [1951] to believe in a distinction (even a rough distinction) between “core” and “non-core” elements associated with a concept, then one already has a theoretically significant notion of conceptual mastery (at least for whichever concepts admit of the core vs. non-core distinction). To resist motivation (a) is to reject the tremendously successful arguments of Burge (1979, 1982, 1986), the conclusions of which have since become philosophical orthodoxy. To resist motivation (b) is to deny ALL BACHELORS ARE MALE any special status, vis-a-vis BACHELOR, above and beyond JAVIER IS A BACHELOR. Few are willing to pay either price. If they do not, they deny the skeptical view.

4.2.5 Four Theories of Conceptual Mastery

I’ve already laid out the distinction between core and non-core elements relative to a concept. The elements might be propositions, inferences, rules, or something else. With the core/non-core distinction in place, a view naturally suggests itself. Fully understanding a concept, i.e. having mastery of that concept, entails having some type of “grip” on that core. This much leaves room for variation. Theories of conceptual mastery can vary according to what they take the grip to consist in. I consider four views.

1. the belief view
2. the inference view
3. the intuition view
4. the meaning postulate view

I plan to argue against the first three views (belief, inference, and intuition), and in favor of my own meaning postulate view. But first let's get the views on the table.

According to the belief view of conceptual mastery, an agent has mastery of concept *C* if and only if he/she believes all of *C*'s core propositions. On this view, the elements of the core are propositions. To grip the core elements is to believe them. According to the inference view of conceptual mastery, an agent has mastery of concept *C* if and only if he/she is disposed to infer in accordance with all of *C*'s core inferences. On this view, the elements of the core are inferences. To grip the core elements is to be disposed to infer in accordance with them. Proponents of the inference view include Peacocke (1992, 1999).¹⁰

The belief and inference view are similar. Let's consider an example of each. Suppose that "all bachelors are male" and "all bachelors are unmarried" are the core truths for BACHELOR. The belief view claims that an agent has mastery of BACHELOR if and only if she believes both that all bachelors are male and that

¹⁰Peacocke (1992) offers a theory of what he calls "concept possession". However, Peacocke does not mean what I mean by 'concept possession' or 'possess(es) a concept'. Peacocke (29-30) states that a thinker can consider propositions, and entertain thoughts, that contain the concept while failing to possess the concept. On my use of 'possess a concept', to possess a concept just is to be able to entertain thoughts or propositions that contain the concept. Thus Peacocke does not mean what I do by 'possess a concept'. His theory of "concept possession" is a theory of a more demanding relation: what I call 'conceptual mastery'.

all bachelors are unmarried. The inference view claims that there are two core inferences associated with BACHELOR. We might express them as follows:

$\frac{x \text{ is a bachelor}}{x \text{ is male}}$	$\frac{x \text{ is a bachelor}}{x \text{ is not married}}$
---	--

The variable ‘x’ represents the fact that the agent is willing to infer from “x is a bachelor” to “x is male” for any x.

According to the intuition view of conceptual mastery, an agent has mastery of concept C when he/she has intuitions that the core propositions are true (Bealer (2002a): 221-230¹¹). On this view, the elements of the core are propositions. To grip the core elements is to intuit their truth. The intuition view relies on a relation of intuition between an agent and a proposition. There is a rich philosophical literature on the nature of intuition. I wish to be as non-committal as possible regarding those debates. Some maintain that intuition is its own sui generis mental attitude (Bealer (2002a,b)). Others claim that intuition is a form of belief (Sosa (1996, 1998), Williamson (2004, 2008)). Even if intuition is a form of belief, the intuition view will, most likely, not reduce to the belief view, because the belief view claims that conceptual mastery requires belief simpliciter, whereas the intuition view claims that conceptual mastery requires a particular form or type of belief (whichever type of belief intuitions are).

According to the meaning postulate view of conceptual mastery, an agent has mastery of concept C if and only if he/she takes all the core elements to govern the use of C. The view is liberal regarding the nature of core elements. To grip a

¹¹Bealer calls mastery of a concept “determinate possession” of the concept. His theory is quite a bit more nuanced, and quite a bit more complicated, than the simplistic intuition view that I consider. I will not explain his view in full here. However, his theory of conceptual mastery is a form of the intuition view, shares its same general shape, and has most of the same advantages and disadvantages.

core element is to take it to govern the use of the concept. In section 4.5.2 (“Taking to Govern Use”) I explain what it is to take something to govern the use of the concept.

There are a variety of dimensions of variations even among these views. One could claim that mastery requires only a grip on some, perhaps a majority, of the elements in the core. Or one might hybridize views, claiming that mastery is a matter of believing some propositions, inferring in accordance with others, and intuiting the truth of a third class. For simplicity’s sake, I ignore these possibilities. I believe they do not significantly alter the dialectic.

4.3 Deviant Masters: Against the Necessity of Belief, Inference, or Intuition for Conceptual Mastery

4.3.1 Overview

In section 4.3 I argue against the necessity of belief in, inference in accord with, or intuition of the truth of, core elements for conceptual mastery. The argument revolves around the possibility of non-standard (“deviant”) masters of the concepts in question. The deviant masters fully understand the concept but neglect to believe, infer, or intuit in accordance with core elements.

The inference view will be my representative stalking horse. In sections 4.3.2-4.3.5 I discuss only the inference view. The belief and intuition views can be defeated by arguments exactly analogous to those I discuss. The modifications required are minimal and obvious.

4.3.2 The Argument

In this section, I’ll argue against the inference view. The basic problem with the view is that an agent can fully understand a concept (i.e. have conceptual mas-

tery) without being disposed to infer in accordance with core inference patterns. Being disposed to infer in accordance with a concept's core inferences is not necessary for mastery of the concept.

Consider the concept *BOCHE*, a racial pejorative applied to Germans. *BOCHE* has two core inference rules.

$\frac{x \text{ is German}}{x \text{ is a boche}}$	$\frac{x \text{ is a boche}}{x \text{ is cruel}}$
--	---

If a thinker is disposed to make both these inferences, then she is disposed to infer from "x is German" to "x is cruel". According to the inference view, an agent who has mastered *BOCHE* must be disposed to infer from "x is German" to "x is cruel".

Here's the counterexample. Consider Agnes, who at one point in time fully understood *BOCHE*. She was disposed to infer, and often did infer, in the manner just described. However, during her travels in Germany, Agnes dined, lived, and fraternized with many Germans. She made German friends. As a result of her experiences, Agnes ceased to be disposed to infer from a subject's German-hood to his or her cruelty. Agnes stopped believing that all Germans are cruel. However, Agnes still realizes that the inference patterns above play a central role in relation to the meaning of *BOCHE* and in the use of 'boche'. This thought experiment yields the following argument against the inference view.

(P1) If the inference view is true, then Agnes has lost her conceptual mastery of *BOCHE*.

(P2) Agnes has not lost her conceptual mastery of *BOCHE*.

(C) Therefore: the inference view is false.

Premise (P1) follows from the statement of the inference view in combination with the description of the thought experiment. Premise (P2) is tremendously plausible. Agnes understands BOCHE at least as well as, and perhaps better than, her unenlightened co-linguals.¹² Agnes still recognizes the inference patterns as rules that govern the use of BOCHE. She simply refuses to obey them. The argument is valid. The conclusion follows: the inference view is false, because being disposed to infer in accordance with core inferences is not necessary for conceptual mastery.

4.3.3 Response 1: Agnes remains disposed to infer

Proponents of the inference view might respond by denying premise (P2), claiming that Agnes remains disposed to infer from “x is German” to “x is cruel”. This disposition is not manifested because it is masked by her belief that not all Germans are cruel.

Two points should be made against this response. First, the move is unmotivated and ad hoc. *Prima facie*, Agnes lacks the disposition. To maintain otherwise one must posit reasons in favor of her having a masked disposition. Proponents of the inference view have offered no such reason, other than that a masked disposition supports their favored theory. Second, a version of the Agnes case can be cooked up in which Agnes clearly lacks the disposition. Whatever the criteria are for having the disposition to infer, we can imagine a case in which Agnes fails to meet those criteria yet has explicit knowledge of the inference rules governing the use of BOCHE. We can generate such a case as long as it’s possible to know what the inference rules governing a concept are without being disposed to obey them. This is possible, so we’ll be able to generate counterexamples to the inference

¹²Williamson (2009): 141 seems to agree.

view, whatever opponents claim about what it is to have a disposition to infer.

4.3.4 Response 2: BOCHE is not a concept

BOCHE is an odd concept. Its inference rules are not truth-preserving. Someone who follows its constitutive inference rules will be led, quite rapidly, into error. (Few Germans are cruel). Such concepts can be called 'defective'. Proponents of inference-style views have sometimes claimed that there are no concepts like BOCHE (e.g. Horwich (2010): 203-4, Peacocke (1992): 21, 171-5, Boghossian (2001)). If there is and cannot be any BOCHE concept, my counterexample fails.

I'll make four points in response to the claim that there are no concepts such as BOCHE. First, the move is (once again) ad hoc. Denial of BOCHE and its ilk is motivated purely by the desire to save the theory from counterexample. Second, rival views, including my preferred meaning postulate view, can accept the existence of such defective concepts. Other things being equal, this counts in favor of the meaning postulate view and against the inference view. (Similar points apply against the belief and intuition views). Third, the move is implausible. *Prima facie*, BOCHE is a concept. People think thoughts containing it, including THAT JERK IS BOCHE. The behavior of those who use BOCHE is repugnant precisely because they use a racist concept and express racist thoughts, not because they use no concept and fail to express any thought.¹³ Fourth, and perhaps most importantly, the argument can be made without using a defective concept such as BOCHE. In the next section, I'll offer a version of the argument using the concept of the material conditional and its core inference *modus ponens*, often taken to be the paradigm on which the inference view builds its theory of all concepts.

¹³Boghossian (2003): 242-3 and Williamson (2009): 140 agree.

4.3.5 A Counterexample Without a Defective Concept

We all encounter the concept of the material conditional \rightarrow in logic class, and think with the concept long before that. In this section, I'll argue that one can master this concept while denying that *modus ponens*, a core inference for \rightarrow , is a valid form of inference. My case is taken directly from Williamson (2008)'s case of the deviant logician (chapter 4, pp. 92-4). These are the two core inferences associated with the concept \rightarrow :

$$\begin{array}{c} P \rightarrow Q \\ P \\ \hline Q \end{array} \qquad \begin{array}{c} P \\ \hline R \rightarrow P \end{array}$$

Consider the deviant logician Professor X. Professor X is a distinguished professor of logic, world-renowned with many seminal publications to his name. If anyone is an expert on *modus ponens*, and on the concept \rightarrow , Professor X is. Professor X, for highly theoretical reasons, denies the validity of *modus ponens*. Perhaps he has been convinced by the alleged counterexamples of McGee (1985). Professor X is wrong to deny its validity, but he has well-considered reasons. This thought experiment yields the following argument against the inference view.

(P1) If the inference view is true, then Professor X does not have mastery of the concept \rightarrow .

(P2) Professor X does have mastery of \rightarrow .

(C) Therefore: the inference view is false.

Premise (P1) follows from the statements of the inference view in combination with the description of the thought experiment. Premise (P2) is tremendously

plausible. If anyone has mastery of arrow, Professor X does. He is the one of, if not the, world's foremost experts on the conditional. The argument is valid. The conclusion follows: the inference view is false. Importantly, this argument does not rely on a defective concept such as BOCHE. The inference rules associated with \rightarrow are truth-preserving. Professor X wrongly denies the validity of *modus ponens*.

4.3.6 Extending the argument against the belief and intuition views

Arguments analogous to those given above can be used to defeat the belief and intuition views. The modifications required to counterexample these affiliated views are minimal. I won't go through the details.

The intuition view fares slightly better than the belief or inference views. The intuition view allows that a thinker might fail to believe a core proposition, or be disposed to infer in accordance with a core inference, yet still master the concept nonetheless, as long as the thinker intuited, or somehow otherwise felt, that the core propositions were true. However, the case of Agnes and BOCHE still defeats the intuition view. She has no intuitions that all Germans are cruel, yet masters the concept nonetheless.

4.4 The Special Status Problem: Against the Sufficiency of Belief, Inference, or Intuition for Conceptual Mastery

4.4.1 Overview of the Special Status Problem

In this section I'll argue that the simple belief, inference, and intuition view cannot work. Believing, inferring, or intuiting are not, by themselves, sufficient for mastery. Instead, a thinker must take the core elements (propositions, inferences, etc.) to have some special status. For example, a thinker might believe core

propositions because they are core or otherwise central to the concept. The initial failure of the simple belief, inference, and intuition views lends itself to an obvious fix. However, the fix must take a certain form in order to succeed. I argue that some existing ways of coping with the problem do not work.

4.4.2 Not Merely Belief

Consider two agents who believe all of BACHELOR's core propositions. Sofia and Felipe both believe that all bachelors are male and that all bachelors are unmarried. Sofia believes these propositions because she takes them to be central to the meaning of 'bachelor' and the concept BACHELOR. Felipe, on the other hand, believes that all bachelors are male, but that this is a matter of happenstance. Felipe is open to the possibility of female bachelors, though he does not believe that any currently exist. Felipe's attitude toward ALL BACHELORS ARE MALE is similar to our attitude toward THERE ARE NO BACHELORS ON MARS.¹⁴ It's true that there are no bachelors on Mars. But that has nothing in particular to do with the meaning of 'bachelor' or 'Mars' and no particular salience vis-a-vis BACHELOR or MARS. Sofia has conceptual mastery of BACHELOR. Felipe does not. This counterexamples the belief view. Similar counterexamples can be given to the inference and intuition views.

(P1) Felipe believes all of BACHELOR's core propositions.

(P2) If the belief view is true, then if Felipe believes all of BACHELOR's core

¹⁴One might take this case as motivation to move toward a view on which the core propositions are all necessary in form. The core proposition is not "all bachelors are male" but rather "necessarily, all bachelors are male". This view does not work. There is still a difference between someone who takes "necessarily, all bachelors are male" to have special status with respect to, or to be "part of the meaning" of, BACHELOR and someone who does not. For example, y might take "necessarily, all bachelors are male" to be true because the actual world is the only possible world and actually, all bachelors are male. But x could believe that maleness has nothing to do with the meaning, or nature, of bachelors.

propositions, he has conceptual mastery of BACHELOR.

(P3) Felipe does not have conceptual mastery of BACHELOR.

(C) Therefore: the belief view is false.

Premise (P1) is a stipulation of the case. Whatever the core propositions are, suppose Felipe believes them. (P2) follows from the statement of the belief view. (P3) is very plausible. Felipe is open to the possibility of female bachelors. Therefore he does not have mastery of BACHELOR. The argument is valid. The conclusion follows: the belief view is false.

4.4.3 Sophisticated Views

One lesson of these examples is that to fully understand, and to have conceptual mastery, one must, in some sense, take the core truths (inferences, rules) to have some special status. This point has been recognized before (Peacocke (1992), Greenberg (MS): 16fn.17, Greenberg (2014): 8fn.6). According to the inference view of Peacocke (1992), an agent who makes the appropriate inferences need not fully understand the concept (in Peacocke's terminology, 'possess the concept' (cf. fn. 11). To fully understand, the agent must find those inferences *primitively compelling*. Peacocke correctly recognizes the importance of the agent's taking some special stance toward the core inferences. Peacocke's version of the special stance is finding the inference primitively compelling.

Peacocke's stratagem can be adopted by any of the belief, inference, or intuition views. It can be implemented in a variety of ways. That the thinker find the inferences primitively compelling is one way. One might also maintain that the thinker infer, or believe, because she takes the inferences (or propositions) to be core. Proponents of the intuition view could claim that the agent's intuit that

the propositions are core, or that taking them to be core plays some role in the creation of the intuition. There are a variety of options. I can't and won't canvass them all. Let's call any version of the belief, inference, or intuition view that adopts some method of requiring that beliefs in, dispositions to infer in accordance with, or intuitions with respect to core elements a special status a *sophisticated belief/inference/intuition view*.

4.4.4 The All and Only Problem

Taking a special attitude toward core elements (e.g. finding primitively compelling) is still not sufficient for conceptual mastery. The special attitude must be taken toward *all and only* the core elements. An agent who, in effect, builds extra elements into her conception of the concept, loses her mastery of the concept, despite the fact that she bears the special attitude to every element of the core.

Consider Peacocke's theory. Suppose Malik finds the inferences from "x is a bachelor" to "x is male" and from "x is a bachelor" to "x is unmarried" primitively compelling. Suppose that these inferences completely constitute BACHELOR's core. According to the sophisticated inference view endorsed by Peacocke, Malik has mastery of BACHELOR. But, unfortunately, Malik also finds the inferences from "x is a bachelor" to "x always wears a baseball cap", "x never cleans his apartment", "x plays too many video games", and "x drinks the blood of cats" primitively compelling. Malik has an absolutely insane conception of what bachelors are like, and he takes these inferences to be part of the meaning of BACHELOR. Malik lacks mastery of the concept BACHELOR.

The obvious move is to modify Peacocke's view accordingly. On such a modified view, a thinker has conceptual mastery of C if and only if she is disposed to infer in accordance with C's core inferences and finds all and only those in-

ferences primitively compelling. But this view is crazy. It entails that a thinker can't fully understand (master) more than one concept at a time. If an agent is to master more than one concept, she'll need to find the core inferences of at least two concepts primitively compelling. So she can't find all and only the inferences associated with C primitively compelling.¹⁵ Thus she does not have mastery of C.

I see no way around this problem while sticking with the notion of "primitively compelling inference". Analogue notions of "primitive belief" and "primitive intuition" (whatever those may be) will face the same problem. The deep problem lies in the logical form of the primitively compelling relation. It is a two-place relation: finds-primitively-compelling(thinker T, inference I). The problem can only be solved by a three-place relation. For example, one might use the "finds primitively compelling in virtue of concept C" relation: finds-primitively-compelling-in-virtue-of(thinker T, inference I, concept C). With this relation, a thinker could find some inferences primitively compelling because of the role C₁ plays in the inference, while finding other inferences primitively compelling because of the role C₂ plays. With a 3-place relation, concepts cease stepping on each other's toes by preventing the thinker from mastery of the other.

On further thought, we can motivate the need for a 3-place relation without the problem of concepts preventing mastery of other concepts. Suppose Felipe finds the inferences from "x is a bachelor" to "x is male" and "x is unmarried" primitively compelling. (Any two-place special attitude will work for the example). But Felipe finds the inference from "x is a bachelor" to "x is male" primitively compelling because of the role that MALE plays in the inference. Felipe

¹⁵I make the trivial assumption that a thinker can master two concepts with non-identical cores.

accords this inference no special status vis-a-vis BACHELOR. Instead, Felipe has a quirky conception of MALE, according to which he finds every inference of the form “x is F” to “x is male” primitively compelling. Without a 3-place relation with the concept itself as a relata, a thinker could take the special attitude toward all and only the concepts core elements, but lack conceptual mastery nonetheless.

Summarizing, we’ve covered two distinct reasons for moving from a 2-place special attitude toward core elements to a 3-place relation. First, a 2-place relation does not allow a thinker to adopt the special attitude toward all and only the core elements of C while simultaneously having mastery of more than one concept. A 3-place relation does. Second, a thinker can adopt a special 2-place attitude to all and only the core elements of a concept C while failing to master that concept. The Felipe-BACHELOR demonstrates this point. Instead, the thinker must accord the core elements some special status with respect to C. Again, a 3-place relation yields the correct results.

4.4.5 Moving On

I don’t think the special status problem is fatal for any of the belief, inference, or intuition view. Something more than mere belief, inference, or intuition must be appealed to, but the theories can be modified appropriately. Peacocke (1992) recognized this point. The thinker must adopt a special attitude toward the core elements. But not even this is enough. The special attitude must be taken to core elements *vis-a-vis the concept*. The concept itself must play a role in the special attitude, whatever it is.

The belief, inference, and intuition views have problems. Neither belief, nor inference, nor intuition is either necessary or sufficient for conceptual mastery. The views must be modified to include a 3-place attitude the thinker can take

toward a core element and a concept.

4.5 The Meaning Postulate View

4.5.1 From the Failures of Rivals to the Meaning Postulate View

In this section, I offer my own positive view of what it is required to have mastery of a concept. The view is motivated by the two failures of the inference, belief, and intuition views. These views first failed by neglecting the possibility of masters of concept C who failed to believe, intuit, or be disposed to infer in accordance with core elements. Agnes has mastery of *BOCHE* without believing that all Germans are cruel. Neither believing in, inferring in accordance with, or intuiting the truth of, core elements is necessary for conceptual mastery. The second failure was ignoring the need for thinkers to adopt some special attitude toward core elements. Mastery of a concept C requires more than merely believing C's core propositions. One must believe them, and in some sense, take the propositions to have some important status vis-a-vis C. Neither belief in, inferring in accordance with, or intuiting the truth of, core elements is sufficient for conceptual mastery.

The inference, belief, and intuition views are overly demanding in one sense and not demanding enough in another. They're overly demanding by requiring conceptual masters to actually believe, be disposed to infer, or intuit. They're not demanding enough because they don't require thinkers to accord core elements a special status. The meaning postulate view jettisons the overly demanding aspect of these views and adds the feature they neglect. According to the meaning postulate view, conceptual mastery is purely a matter of taking core elements to have a special status with respect to the concept. That's it. Taking the core elements

to have a special status vis-a-vis C , or to govern the meaning of C , is compatible with not believing those core elements, inferring according to them, or intuiting their truth.

- **The Meaning Postulate View of Conceptual Mastery:** If an agent has mastery of concept C , then he/she takes all and only the elements in C 's core to govern the use of C .

The meaning postulate view is agnostic about the nature of the elements in a concept's core. The elements could be propositions, patterns of inference, mental transitions, rules, or something else altogether. In semantics, meaning postulates are axioms laid down to help establish the meaning of a term. I think of core elements as playing a similar role. The core elements help to give the concept its meaning. A different core would yield a different concept.¹⁶ The view is so titled because the "meaning postulate view" (of conceptual mastery) because to have conceptual mastery requires taking the core elements to play something like the role of meaning postulates. (A more accurate, though less elegant, title might be the "take to be a meaning postulate view").

4.5.2 Taking to Govern Use

To explain the meaning postulate view, I must explicate what it is for an agent to take a core element to govern the use of C . For ease of exposition, I'll speak of core elements as *rules*. I intend to construe 'rule' as broadly as possible. On this

¹⁶One might wonder whether two concepts could have the same core. Probably not, for trivial reasons. C_1 's core will likely mention C_1 , not C_2 . C_2 's core will mention C_2 , not C_1 . So if C_1 and C_2 are distinct they have distinct cores. The more interesting question is whether there could be two distinct concepts with homomorphic cores, in which each element of the core of C_1 could be mapped to an element of C_2 's core by swapping occurrences of C_1 for occurrences of C_2 , and vice-versa. On this question I am officially agnostic.

terminology, inference patterns, propositions, and mental transitions all count as rules.

I will not offer a reductive theory of the “taking” relation. One should not view the taking relation in too heavy-weight a manner. A thinker need not have the concepts CONCEPT, GOVERN THE USE, MEANING POSTULATE, or MEANING CONSTITUTING RULE in order to take a core element to govern the use of the concept. Whether the thinker takes a rule to be a core rule governing use of the concept is determined by idealization from the psychology of the agent. The agent’s past usage and present dispositions should be taken into account. The “takes to govern the use” relation is psychologically real. Whether an agent takes a core proposition to govern is a genuine fact about the agent’s psychology. It is not merely a “stance” that theorists take toward the agent. The taking will, in general, have ramifications on the thinker’s behavior. However, the best way to determine whether a thinker takes a proposition to govern the use (rather than, say, merely taking the proposition to be true) will often be by testing her behavior in counterfactual situations, idealized in various ways.

There are many ways to take a given proposition to govern the use of a concept. Consider the proposition ALL BACHELORS ARE MALE. Most straightforwardly, a sophisticated thinker might directly believe that ALL BACHELORS ARE MALE governs the use of BACHELOR. Professor X’s taking of *modus ponens* to govern the use of the material conditional seems to have something like this form. More commonly, a thinker’s taking ALL BACHELORS ARE MALE to govern the use of BACHELOR will be less sophisticated. She’ll be confounded or perplexed when someone claims that some bachelors are not male. This perplexation will have a different feel than the perplexation she’ll experience when someone claims that

several bachelors live on Mars. She'll predict the interlocutor who claims that bachelors live on Mars to be misinformed. Perhaps he or she heard that the Mars rover Opportunity is a manned mission involving a crew. In contrast, she'll tend to think that the interlocutor who claims that some bachelors are not male misunderstands how to use BACHELOR and what it is to be a bachelor.

Conceivability provides a useful heuristic for gauging what an agent takes to govern the use of a concept. A proposition is *conceivable* for an agent when he or she finds the proposition possible. A proposition is *inconceivable* when she finds the proposition to be impossible. In general, an agent who takes a proposition P to govern the use of a concept will find P inconceivable. For example, someone who takes ALL BACHELORS ARE MALE to govern the use of BACHELOR will tend to find SOME BACHELORS ARE NOT MALE impossible. Someone who does not take it to govern the use of BACHELOR will find such a situation conceivable. This heuristic yields the correct verdict in the case of Sofia and Felipe (section 4.4.2). Both Sofia and Felipe believe the proposition ALL BACHELORS ARE MALE. Sofia, who has mastery of BACHELOR, finds SOME BACHELORS ARE NOT MALE inconceivable. Felipe, who lacks mastery of BACHELOR, finds SOME BACHELORS ARE NOT MALE conceivable. The heuristic does not work well for deviant masters. We can flesh out the BOCHE/Agnes case (4.3.2) so that Agnes finds SOME BOCHES ARE NOT CRUEL conceivable, despite the fact that she takes ALL BOCHES ARE CRUEL to govern the use of BOCHE.¹⁷

¹⁷There are a variety of ways to precisify the BOCHE/Agnes case. Not all of them yield this result. In the case as originally described, it's not clear what Agnes's attitudes toward BOCHE are. There are many possibilities. Perhaps she thinks that no one is a boche. Perhaps she thinks that all Germans are boches but not all boches are cruel. Perhaps she thinks that only some Germans are boches, all of which are cruel. Or perhaps she thinks that some Germans are boches, and some boches are cruel. What Agnes finds conceivable will depend on how we precisify the case. All I claim is that on one reasonable way of precisifying the case the conceivability heuristic will not yield the correct verdict regarding Agnes's mastery of BOCHE.

In sum, I claim that there is a psychologically real difference between an agent who takes ALL BACHELORS ARE MALE to be true because it bears a special status with respect to the use of BACHELOR and the otherwise identical agent who takes ALL BACHELORS ARE MALE to merely happen to be true. In my terminology, the first takes ALL BACHELORS ARE MALE to govern the use of BACHELOR. The second does not, and thereby does not have mastery of BACHELOR. Furthermore, this distinction applies even to agents that do not have sophisticated concepts such as CONCEPT, MEANING, or GOVERN THE USE.

4.5.3 Governing Use

In this section, I'll explain the sense in which a core element, or rule, governs the use of a concept. I explained earlier (section 4.5.1) that core elements play something like the role of meaning postulates in semantics. Meaning postulates help give a linguistic term its meaning. Similarly, core elements help give a concept its meaning. The core elements, or rules, are guidelines laid down for the use of the concept. They have normative force, indicating how one should and should not use the concept. In this sense they govern the use of the concept.

The possibility of defective concepts poses a special problem for this approach to core elements as use-governors. I've claimed that conceptual mastery of concept C entails taking C's core elements to govern the use of C. However, some concepts, such as BOCHE, have core inferences that do not preserve truth. (Alternatively, one could take BOCHE to have core propositions the conjunction of which is false). An agent who realizes that a concept's core elements do not preserve truth should in fact cease thinking (and inferring) in accordance with that element. Agnes is in that situation. She recognizes that inferring from "x is German" to "x is cruel" is a bad inference, despite the fact that's BOCHE's core el-

ements suggest the inference. If core inferences need not preserve truth, in what sense do they “govern the use” of the concept?

I’d like to use an analogy with the laws of a nation to explain the sense in which core elements can govern the use of a concept while failing to yield truth. Suppose that the law, and particular laws, have certain goals, purposes, or functions: to give citizens a better life, to promote equality, to prevent injustice, to prohibit immoral action, etc. (That law, or laws, do have goals or functions in this way, that are not merely the goals of law-makers, is controversial. But the position is certainly coherent. Coherence is all my analogy requires.)¹⁸ Sometimes the law achieves these goals. Sometimes it does not. When a law does not achieve its goal, it remains a law nonetheless and retains its legal normative force. The fact that the law does not, e.g., promote equality, does not thereby render the law no longer legally binding. Those who disobey the law still violate a legal norm. If the law is immoral, then violating the legal norm may be the only way to obey a moral norm.

Rules of use have a certain function as well. They aim, broadly speaking, toward truth. They can aim at truth itself, or at the preservation of truth. (The former is more appropriate to core propositions, the latter to core inferences.) Rules may have other functions as well. But rules of use, much like the laws of a nation, can fail to achieve their aims. The core rules of defective concepts such as *BOCHE* fail in exactly this way. They fail to meet their truth-oriented functional goal. But nonetheless the rules remain rules that govern the use of *BOCHE*. They retain normative force. In some cases, the only way to obey the epistemic norm of believing truth will be to flout the meaning norm governing the use of a concept

¹⁸For discussion of laws as functional, or as having a purpose, cf. Greenberg (2011): 84-88).

such as *BOCHE*. Agnes recognizes something like this fact about her situation. She elects to obey the epistemic norm at the cost of flouting the use norm.

In sum, there is a clear sense in which a rule can generate a norm that governs the use of a concept. In the case of a defective concept, this norm of use might conflict with the epistemic norms of truth and truth-preservation.

4.6 Three Objections

4.6.1 Overview of the Objections

In this section, I consider three objections to the meaning postulate view. The first is that the theory appeals to a convoluted and under-explained notion of “taking a rule to govern the use of the concept”. The second objection is more direct. It claims that an agent can have conceptual mastery without taking core propositions to govern the use. Thus taking to govern use is not necessary for conceptual mastery. The third objection maintains that taking to govern use is not sufficient for conceptual mastery. Mastery of some concepts requires certain abilities.

4.6.2 Objection 1: Against the notion of “taking a rule to govern the use”

Opponents might object to the notion of “taking a rule to govern the use” on which the meaning postulate view relies. Opponents might maintain that the notion is under-explained. Or they might maintain that the notion is too close the target notions of conceptual mastery and “full understanding” of a concept.

Perhaps the notion of “taking a rule to govern the use” is under-explained. I’ve done some work to get the reader on to the idea. One reason for under-explanation is that I’ve tried to be relatively agnostic about the nature of core rules. They could be propositions, rules of inference, or something else altogether.

I'm inclined to think that different types of concepts will require different types of core elements. My agnosticism has a theoretical basis. Clearly, there is further work to be done in exploring the notion of "taking a rule to govern the use of a concept". But I take myself to have made some headway.

I admit that "taking a rule to govern the use" is more similar to "fully understanding, or mastering, the concept" than believing or inferring. But I think that this is a step in the right direction. The failures of the belief and inference views are partly due to their overly reductive ambitions. To achieve a successful theory of conceptual mastery, we must appeal to notions that are closer in nature to understanding itself. "Taking to govern the use" is closer to understanding. That's part of the reason the meaning postulate view succeeds where its rivals fail.

My main reply to the objection is the "it's everyone's problem" response. In section 4.4.2 I argued that the simple versions of the belief, inference, and intuition view do not work. There's a significant difference in understanding between someone who merely infers and someone who infers because that inference governs the use of the concept, or plays some other special role vis-a-vis the concept. The sophisticated versions of these views accommodate the point. But any view that accommodates the point must appeal to something like "taking a rule to govern the use". Peacocke (1992) does so with the notion of a "primitively compelling" inference. Thus, every theory of conceptual mastery (or at least all the options I've considered) will appeal to something like "taking a rule to govern the use". Perhaps the notion is under-explained. Or perhaps it is close to the target notion of full understanding of a concept. But everyone must appeal to some notion in the vicinity. So it's everyone's problem. (I admit that some "notions in the vicinity" might be more or less problematic than others. Perhaps "primi-

tively compelling inference” is less problematic than “taking to govern the use”. I’m not sure. If it is, I’d like to see the arguments.)

4.6.3 Objection 2: Mastery Without Taking to Govern Use

This objection maintains that a thinker can have conceptual mastery of a concept without taking a core rule to govern the use. Thus the meaning postulate view is false, because taking to govern the use is not necessary for conceptual mastery. Here’s the counterexample. Consider Professor Y, a philosopher of mind and language. She works on concepts. She had read this paper. She detests it. Professor Y believes that there is no distinction between core and non-core rules, no such thing as a rule that “governs the use” of the concept. Professor Y’s theoretical commitments lead her to deny every claim of the form “Rule R governs the use of concept C”. This thought experiment leads to the following argument against the meaning postulate view.

(P1) Professor Y has conceptual mastery of many concepts.

(P2) Professor Y does not take any rules to govern the use of those concepts.

(P3) If (P1) and (P2), the meaning postulate view is false.

(C) Therefore: the meaning postulate view is false.

Premise (P1) is very plausible. Everyone has mastery of at least a few concepts (e.g. AND, HERS, HERE). Furthermore, Professor Y is a world-authority on many concepts, including the concepts CONCEPT, MEANING, TRUTH, and RULE OF USE. If anyone fully understands these concepts, she does. Premise (P2) is quite plausible as well. Professor Y denies that rules govern the use of concepts. Thus she does not take any rules to govern the use of any concept. Premise (P3)

follows from the statement of the meaning postulate view, which maintains that taking rules to govern use is a necessary condition for conceptual mastery.

I deny premise (P2). Professor Y denies all claims of the form “rule R governs the use of concept C”. She believes many statements of the form “rule R does not govern the use of concept C”. I claim that, nonetheless, Professor Y takes many rules to govern the use of many concepts. Consider her behavior. She will look at you blankly when you talk about female bachelors. She won’t understand if you attempt to use the term ‘here’ to refer to a location other than the current one (or the location demonstrated). She’ll claim (perhaps unreflectively) that you’re incorrectly using the concept HERE. Professor Y’s belief that rules do not govern use, and her taking them not to govern use, is compatible with her taking certain rules to govern use. Furthermore, I don’t see how Professor Y could think, or use any concepts at all, if she never took any rule to govern the use of any concept. The rules guide her use of the concepts.

4.6.4 Objection 3: Taking to Govern Use Without Mastery

I have claimed elsewhere (Rabin (2011a)) that mastery of phenomenal experiential concepts such as RED_{ph} requires certain abilities, including the ability to recognize a red sensation as an instance of RED_{ph}.¹⁹ If this is correct, then it seems that someone could take whatever the core rules associated with RED_{ph} (perhaps “red_{ph} is a sensation”) are to govern the use of RED_{ph} whilst being unable to recognize a red_{ph} experience as red_{ph}. Let Jane be a color-blind person who does not

¹⁹RED_{ph} is the phenomenal, experiential, concept of red. RED_{ph} applies to experiences with a certain qualitative character (you know the one). Neither light-waves nor objects can be red_{ph}. Tomatoes, fire trucks, and strawberries are red, but they cause red_{ph} sensations. The English term ‘red’ is likely polysemous between, or a mongrel of, the red-concepts of light waves, of surfaces, and of sensations. I introduce a new term, ‘red_{ph}’, and concept RED_{ph}, to lock on to the type of red that only sensations can instantiate.

experience red_{ph} at all. She experiences green_{ph} instead. She cannot recognize a red_{ph} sensation as an instance of RED_{ph} . She's never experienced red_{ph} . But Jane, a philosopher of mind and meaning, is an expert on the semantics of 'red' and on the concept RED_{ph} . She certainly seems to know what the rules governing the use of the concept are. This thought experiment leads to the following argument against the meaning postulate view.

(P1) Jane takes all and only the rules governing the use of RED_{ph} to govern the use of RED_{ph} .

(P2) If (P1), then if the meaning postulate view is true, Jane has conceptual mastery of RED_{ph} .

(P3) Jane is unable to recognize a red_{ph} sensation as an instance of RED_{ph} .

(P4) If (P3), then Jane does not have conceptual mastery of RED_{ph} .

(C) Therefore: the meaning postulate view is false.

Premise (P1) seems to follow from Jane's philosophical expertise regarding the concept RED_{ph} . (P2) follows from the statement of the meaning postulate view. (P3) is a stipulation about the case. (P4) is a plausible claim about any phenomenal concept, including RED_{ph} . I argue for (P4) in Rabin (2011a).

One simple response would be to accept the argument and weaken the meaning postulate view by claiming that taking all and only a concept's core rules to govern use is only a necessary, but not a sufficient, condition for mastery of the concept. But doing so significantly reduces the interest and the scope of the meaning postulate view.

I prefer to deny premise (P1). The key idea here is to consider the possibility that the ability to recognize red_{ph} experiences as instances of RED_{ph} is itself a core rule. An inference rule is a form of mental transition. The shift from an experience of red_{ph} to a labeling of that experience as RED_{ph} is also a mental transition. This transition could be a core rule governing the use of RED_{ph} . On reflection, this seems exactly correct. There is something about labeling red_{ph} experiences as red_{ph} that is central to the meaning and use of RED_{ph} . Here, we see a broadening of the kinds of elements that can occur in a concept's core.²⁰ Before we considered propositions and inference rules. Now we include mental transitions considered more broadly. (In some sense, perhaps the transition from a red_{ph} experience to a labeling as red_{ph} is a variety of "inference", broadly construed. Not much turns on the nomenclature.)

If it is correct that the mental transition from a red_{ph} experience to a recognition that the experience is ph is a core rule governing the use of RED_{ph} , then premise (P1) is false. The fact that Jane cannot recognize red_{ph} experiences as red_{ph} entails that she fails to take a core rule governing the use of RED_{ph} as such.

As should be expected by now, merely making the transition from experiencing red_{ph} to believing the experience to be red_{ph} is not sufficient to take that transition to be a core rule governing RED_{ph} . A thinker might identify a red_{ph} experience as an instance of RED_{ph} because that experience occurs in the area of the visual field occupied by an image of a fire truck. The thinker knows that fire trucks cause red_{ph} experiences. The thinker must instead categorize the experience on the basis of its phenomenal qualities, and take the fact that objects of this quality are red_{ph} to help govern the use of RED_{ph} .

²⁰Peacocke (1992): 7-8 adopts a similar move.

4.7 Conclusion

For now, this ends my attempt to answer the question “Under what conditions does an agent master, or fully understand, a concept?” Even if the reader does not agree with my positive theory, the meaning postulate view, he or she should find the criticisms of the belief, inference, and intuition views illuminating. The phenomenon of deviant masters and the special status problem must be dealt with by any theory of conceptual mastery.

I conclude that mastering, or fully understanding, a concept is not a matter of merely having certain beliefs, making certain inferences, or intuiting certain truths. Mastering a concept is a matter of recognizing, perhaps implicitly, the rules governing the use of the concept. This should not come as a surprise. Using a concept includes being subject to a normative standard associated with that concept. This normative standard is determined, in large part, by the rules that govern the use of the concept. To understand the concept is to grasp those rules.

Bibliography

Alter, Torin. forthcoming. Social Externalism and the Knowledge Argument. *Mind*.

Ball, Derek. 2009. There are no Phenomenal Concepts. *Mind*, **118**, 935–62.

Ball, Derek. forthcoming. Consciousness and Conceptual Mastery. *Mind*.

Balog, Katalin. 1999. Conceivability, Possibility, and the Mind-Body Problem. *The Philosophical Review*, **108**, 497–528.

Balog, Katalin. 2012. Acquaintance and the Mind Body Problem. *Pages 16–43 of: Hill, Christopher, & Gozzano, Simone (eds), New Perspectives on Type Identity: The Mental and the Physical*. Cambridge University Press.

Bealer, George. 2002a. Intuition and the Autonomy of Philosophy. *In: DePaul & Ramsey (1998)*.

Bealer, George. 2002b. Modal Epistemology and the Rationalist Renaissance. *In: Hawthorne & Gendler (2002)*.

Block, Ned. 1995. On A Confusion About A Function of Consciousness. *The Behavioral and Brain Sciences*, **18**, 227–47.

Block, Ned, & Stalnaker, Robert. 1999. Conceptual Analysis, Dualism, and the Explanatory Gap. *Philosophical Review*, **108**, 1–46.

- Boghossian, Paul. 2001. How are Objective Epistemic Reasons Possible? *Philosophical Studies*, **106**, 1–40.
- Boghossian, Paul. 2003. Blind Reasoning. *Aristotelian Society Supplementary*, **77**, 225–248.
- Braddon-Mitchell, David. 2003. Qualia and Analytical Conditionals. *Journal of Philosophy*, **100**, 111–135.
- Burge, Tyler. 1979. Individualism and the Mental. *Midwest Studies in Philosophy*, **IV**, 73–122.
- Burge, Tyler. 1982. Other Bodies. *Pages 97–120 of: Woodfield, A. (ed), Thought and Object: Essays on Intentionality*. Clarendon Press.
- Burge, Tyler. 1986. Intellectual Norms and Foundations of Mind. *Journal of Philosophy*, **83**, 697–720.
- Burge, Tyler. 1990. Frege On Sense and Linguistic Meaning. *Pages 30–60 of: Bell, D., & Cooper, N. (eds), The Analytic Tradition*.
- Byrne, Alex. 2002. Something About Mary. *Grazer Philosophische Studien*, **63 (Essays in Honour of the Philosophy of Terence Horgan)**, 123–40.
- Byrne, Alex. 2006. Review of “There’s Something About Mary”. *Notre Dame Philosophical Reviews*.
- Carnap, Rudolf. 1929. *Der Logisches Aufbau der Welt*. Berlin: Schlachtensee Weltkreis-Verlag.
- Chalmers, David. 1996a. *The Conscious Mind*. Oxford: Oxford University Press.

- Chalmers, David. 1996b. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chalmers, David. 1999. Materialism and the Metaphysics of Modality. *Philosophy and Phenomenological Research*, **59**, 473–93.
- Chalmers, David. 2002. Does Conceivability Entail Possibility? *In: Hawthorne & Gendler (2002)*.
- Chalmers, David. 2003a. Consciousness and Its Place in Nature. *Pages 102–142 of: Stephen Stich, Ted Warfield (ed), The Blackwell Guide to Philosophy of Mind*. Malden, MA: Blackwell.
- Chalmers, David. 2003b. The Content and Epistemology of Phenomenal Belief. *Pages 220–272 of: Smith, Q., & Jokic, A. (eds), Consciousness: New Philosophical Perspectives*. Oxford.
- Chalmers, David. 2006a. The Foundations of Two-Dimensional Semantics. *Pages 55–140 of: Garcia-Carpintero, Manuel, & Macia, Josep (eds), Two-Dimensional Semantics*. Oxford University Press.
- Chalmers, David. 2006b. Two-Dimensional Semantics. *Pages 574–606 of: Lepore, Ernest, & Smith, Barry C. (eds), Oxford Handbook of Philosophy of Language*. New York: Oxford University Press.
- Chalmers, David. 2010. The Two-Dimensional Argument Against Materialism. *Pages 141–191 of: The Character of Consciousness*.
- Chalmers, David. 2012. *Constructing the World*. Oxford: Oxford University Press.
- Chalmers, David. forthcoming. *Constructing the World*. Oxford.

- Chalmers, David, & Jackson, Frank. 2001. Conceptual Analysis and Reductive Explanation. *Philosophical Review*, **110**, 315–61.
- Conee, Earl. 1994. Phenomenal Knowledge. *Australasian Journal of Philosophy*, **72**, 136–50.
- Correia, Fabrice, & Schneider, Benjamin (eds). 2012. *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge University Press.
- DePaul, Michael, & Ramsey, William (eds). 1998. *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Rowman and Littlefield.
- deRosset, Louis. ms. Semantics for the Pure Logic of Ground.
- Dowell, J.L. 2006a. Formulating the Thesis of Physicalism: An Introduction. *Philosophical Studies*, **131**, 1–23.
- Dowell, J.L. 2006b. The Physical: Empirical not Metaphysical. *Philosophical Studies*, **131**, 25–60.
- Evans, Gareth. 1973. The Causal Theory of Names. *Proceedings of the Aristotelian Society*, **47**, 187–208.
- Fine, Kit. 2001. The Question of Realism. *Philosopher's Imprint*, **1**, 1–30.
- Fine, Kit. 2011. The Pure Logic of Ground. *The Review of Symbolic Logic*, **5**, 1–25.
- Fine, Kit. 2012. A Guide to Ground. In: Correia & Schneider (2012).
- Fitch, Frederic. 1963. A Logical Analysis of Some Value Concepts. *The Journal of Symbolic Logic*, **28**, 135–142.

- Frege, Gottlob. 1952/1892a. On Sense and Reference. *In: Geach, Peter, & Black, Max (eds), Translations of the Philosophical Writings of Gottlob Frege.* Oxford: Blackwell.
- Frege, Gottlob. 1952/1892b. The Thought: A Logical Enquiry. *In: Geach, Peter, & Black, Max (eds), Translations of the Philosophical Writings of Gottlob Frege.* Oxford: Blackwell.
- Gill, Mary Louise. 1989. *Aristotle on Substance: The Paradox of Unity.* Princeton, NJ: Princeton University Press.
- Graham, George, Horgan, Terry, & Tienson, John. 2007. Consciousness and Intentionality. *Pages 468–484 of: Velmans, M., & Schnieder, S. (eds), The Blackwell Companion to Consciousness.*
- Greenberg, Mark. 2001. *Thoughts Without Masters: Incomplete Understanding and the Content of Mind.* Ph.D. thesis, Oxford University.
- Greenberg, Mark. 2005. A New Map of Theories of Mental Content. *Nous*, **39**(1), 299–320.
- Greenberg, Mark. 2009. Moral Concepts and Motivation. *Philosophical Perspectives*, **23**, 137–64.
- Greenberg, Mark. 2011. The Standard Pictures and Its Discontents. *Pages 39–106 of: Oxford Studies in Philosophy of Law.*
- Greenberg, Mark. 2014. Meta-Semantics and the Distinction between Concept Possession and Concept Mastery. *Pages ??–?? of: Burgess, A., & Sherman, B. (eds), New Essays in Metasemantics.*

- Greenberg, Mark. MS. *Incomplete Understanding, Deference, and the Content of Thought*.
- Hawthorne, John. 2002. Advice for Physicalists. *Philosophical Studies*, **108**, 17–52.
- Hawthorne, John, & Gendler, Tamar Szabo (eds). 2002. *Conceivability and Possibility*. Oxford University Press.
- Hellie, Benj. 2004. Inexpressible Concepts and the Allure of the Knowledge Argument. *Pages 333–64 of: Peter Ludlow, Yujin Nagasawa, & Stoljar, Daniel (eds), There's Something About Mary*. MIT Press.
- Hill, Christopher. 1997. Imaginability, Conceivability, Possibility, and the Mind-Body Problem. *Philosophical Studies*, **87**, 61–85.
- Horgan, Terry, & Tienson, John. The Phenomenology of Intentionality and the Intentionality of Phenomenology. *Pacific Philosophical Quarterly*, **84**, 1–23.
- Horwich, Paul. 2010. *Truth - Meaning - Reality*. Oxford University Press.
- Jackson, Frank. 1982. Epiphenomenal Qualia. *Philosophical Quarterly*, **32**, 127–136.
- Jackson, Frank. 1986. What Mary Didn't Know. *The Journal of Philosophy*, **83**, 291–95.
- Jackson, Frank. 1993. Armchair Metaphysics. *Pages 23–42 of: Hawthorne, John, & Michael, Michaelis (eds), Philosophy In Mind: The Place of Philosophy in the Study of Mind*. Dordrecht: Kluwer.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.

- Jackson, Frank. 2005. The Case for A Priori Physicalism. *Pages 102–42 of: Nimtz, Christian, & Beckermann, Ansgar (eds), Philosophy-Science-Scientific Philosophy, Main Lectures and Colloquia of GAP 5, Fifth International Congress of the Society for Analytical Philosophy.* Mentis.
- Kaplan, David. 1970. Dthat. *Syntax and Semantics*, **9**, 221–43.
- Kaplan, David. 1989. Demonstratives. *In: Almog, Joseph, Perry, John, & Wettstein, Howard (eds), Themes from Kaplan.* New York: Oxford University Press.
- Kim, Jaegwon. 1982. Psychophysical Supervenience. *Philosophical Studies*, **41**, 51–70.
- Kripke, Saul. 1972. Naming and Necessity. *Pages 253–355, 763–769 of: Davidson, Donald, & Harman, Gilbert (eds), Semantics of Natural Language.* Dordrecht: Reidel. Revised edition published in 1980 as *Naming and Necessity* (Cambridge, MA: Harvard University Press).
- Leuenberger, Stephan. 2008. *Ceteris Absentibus* Physicalism. *Pages 145–70 of: Zimmerman, Dean (ed), Oxford Studies in Metaphysics, volume 4.* Oxford: Oxford University Press.
- Levine, Joseph. 1983. Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, **64**, 354–361.
- Levine, Joseph. 2004. *Purple Haze: The Puzzle of Consciousness.* Oxford University Press.
- Levine, Joseph, & Trogon, Kelly. 2009. The modal status of materialism. *Philosophical Studies*, **145**, 351–362.

- Lewis, David. 1983. New Work for a Theory of Universals. *Australasian Journal of Philosophy*, **61**, 343–77.
- Lewis, David. 1986. *On the Plurality of Worlds*. Oxford: Basil Blackwell.
- Lewis, David. 1988. What Experience Teaches. *Proceedings of the Russellian Society*, **13**, 29–57.
- Loar, Brian. 1990. Phenomenal States. *Philosophical Perspectives*, **4**, 81–108.
- McGee, Vann. 1985. A Counterexample to Modus Ponens. *Journal of Philosophy*, **82**, 462–471.
- McGinn, Colin. 1989. Can We Solve the Mind-Body Problem? *Mind*, **98**, 349–66.
- Melnyk, Andrew. 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge University Press.
- Millikan, Ruth Garrett. 1984. *Language, Thought and Other Biological Categories*. MIT Press.
- Millikan, Ruth Garrett. 1989. Biosemantics. *Journal of Philosophy*, **86**, 281–97.
- Millikan, Ruth Garrett. 1993. *White Queen Psychology and Other Essays for Alice*. MIT Press.
- Nagel, Thomas. 1974. What is It Like to be a Bat? *Philosophical Review*, **83**, 435–50.
- Nemirow, Laurence. 1990. Physicalism and the Cognitive Role of Acquaintance. *Pages 490–99 of: Lycan, William (ed), Mind and Cognition*.
- Nida-Rumelin, Martina. 1996. What Mary couldn't know: belief about phenomenal states. *Pages 219–42 of: Metzinger, T. (ed), Conscious Experience*.

- Nida-Rumelin, Martina. 2009. *Qualia: The Knowledge Argument*.
- Papineau, David. 1993. Physicalism, Consciousness, and the Antipathetic Fallacy. *Australasian Journal of Philosophy*, **71**, 169–183.
- Papineau, David. 1998. Mind the Gap. *Philosophical Perspectives*, **12**, 373–89.
- Papineau, David. 2002. *Thinking About Consciousness*. Oxford: Oxford University Press.
- Peacocke, Christopher. 1992. *A Study of Concepts*. MIT Press.
- Peacocke, Christopher. 1999. *Being Known*. Oxford University Press.
- Peacocke, Christopher. 2008. *Truly Understood*. Oxford University Press.
- Putnam, Hilary. 1962. It ain't necessarily so. *Journal of Philosophy*, **22**, 658–71.
- Putnam, Hilary. 1975. The Meaning of Meaning. *Pages 131–93 of: Gunderson, Keith (ed), Language, Mind and Knowledge*. Minnesota Studies in the Philosophy of Science, vol. 7. Minneapolis: University of Minnesota Press.
- Quine, W. V. O. 1951. Two Dogmas of Empiricism. *Philosophical Review*, **60**, 20–43.
- Rabin, Gabriel. 2011a. Conceptual Mastery and the Knowledge Argument. *Philosophical Studies*, **54**, 125–147.
- Rabin, Gabriel. 2011b. Physicalism. *Australasian Journal of Philosophy*, **89**, 562–66.
- Rabin, Gabriel. MS. *Physicalism and the Idealization of A Priori Entailment*.
- Rabin, Gabriel, & Rabern, Brian. ms. Well Founding Grounding Grounding.

- Rosen, Gideon. 2010. Metaphysical Dependence, Grounding, and Reduction. *In: Hale, Bob, & Hoffman, Aviv (eds), Modality: Metaphysics, Logic, and Epistemology.*
- Russell, Bertrand. 1927. *The Analysis of Matter*. London: Routledge.
- Salmon, Nathan. 1986. *Frege's Puzzle*. Cambridge, MA: MIT Press.
- Sawyer, Sara. 2003. Conceptual Errors and Social Externalism. *The Philosophical Quarterly*, **53**, 265–273.
- Schaffer, Jonathan. 2003. Is there a fundamental level? *Nous*, **37**(3), 498–517.
- Schaffer, Jonathan. 2009. On What Grounds What. *Pages 346–383 of: Chalmers, David, Manley, David, & Wasserman, Ryan (eds), Metametaphysics*. Oxford University Press.
- Schaffer, Jonathan. 2010. Monism: The priority of the whole. *The Philosophical Review*, **119**(1), 31.
- Schiffer, Stephen. 1989. *Remnants of Meaning*. Cambridge, MA: MIT Press.
- Schwarz, Wolfgang. 2007. Conceptual Metaphysics and Modal Metaphysics. *In: Helen Bohnse, Kirstin Dreimann, & Walter, Sven (eds), Selected Papers Contributed to the Sections of GAP.6, Sixth International Congress of the German Society for Analytical Philosophy*. Paderborn: Mentis.
- Seiberg, Nathan. 1996. *Emergent Spacetime*. <http://arxiv.org/abs/hep-th/0601234>, Rapporteur talk at the 23rd Solvay Conference in Physics, December, 2005.

- Smeenk, Christopher. 2014. Einstein's Role in the Creation of Relativistic Cosmology.
- Sosa, Ernest. 1996. Rational Intuition: Bealer on Its Nature and Epistemic Status. *Philosophical Studies*, **8**, 151–162.
- Sosa, Ernest. 1998. Rational Intuition: Bealer on Its Nature and Epistemic Status. *In: DePaul & Ramsey (1998)*.
- Stoljar, Daniel. 1996. Nominalism and Intentionality. *Nous*, **30**, 261–81.
- Stoljar, Daniel. 2001. Two Conceptions of the Physical. *Philosophy and Phenomenological Research*, **LXIII**, 253–281.
- Stoljar, Daniel. 2005. Physicalism and Phenomenal Concepts. *Mind and Language*, **20**, 469–94.
- Stoljar, Daniel. 2009a. Physicalism. *Stanford Encyclopedia of Philosophy*.
- Stoljar, Daniel. 2009b. *Physicalism*. Routledge.
- Strawson, Galen. 1996. *Mental Reality*. MIT Press.
- Tye, Michael. 1995. *Ten Problems of Consciousness*. MIT Press.
- Tye, Michael. 2000. *Color, Consciousness, and Content*. MIT Press.
- Tye, Michael. 2009. *Consciousness Revisited: Materialism without Phenomenal Concepts*. Cambridge: MIT Press.
- Wikforss, Asa. 2001. Social Externalism and Conceptual Errors. *The Philosophical Quarterly*, **51**, 217–231.

- Wikforss, Asa. 2004. Externalism and Incomplete Understanding. *The Philosophical Quarterly*, **54**, 287–294.
- Wikipedia. 2013. *Lake Victoria*.
- Williamson, Timothy. 2004. Philosophical “Intuitions” and Scepticism about Judgement. *Dialectica*, **58**, 109–53.
- Williamson, Timothy. 2008. *The Philosophy of Philosophy*. Oxford: Blackwell.
- Williamson, Timothy. 2009. Reference, Inference, and the Semantics of Pejoratives. *Pages 137–158 of: The Philosophy of David Kaplan*.
- Williamson, Timothy. draft. How Deep is the Distinction between A Priori and A Posteriori Knowledge?
- Wilson, Jessica. 2005. Supervenience-based Formulations of Physicalism. *Nous*, **39**, 426–459.
- Wilson, Jessica. 2006. On Characterizing the Physical. *Philosophical Studies*, **131**, 61–99.
- Worley, Sara. 2006. Physicalism and the Via Negativa. *Philosophical Studies*, **131**, 25–60.
- Wright, Crispin. draft. *The Basic A Priori: Arithmetic as a Case Study*.