



LIBRARY  
Michigan State  
University

This is to certify that the  
dissertation entitled

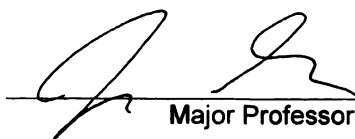
PHYSICALISM AND QUALITATIVE FACTS:  
A CRITIQUE OF FRANK JACKSON

presented by

Noel Boyle

has been accepted towards fulfillment  
of the requirements for the

Ph.D. degree in Philosophy



Major Professor's Signature

8-2-08

Date

**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.  
**MAY BE RECALLED** with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

**PHYSICALISM AND QUALITATIVE FACTS:  
A CRITIQUE OF FRANK JACKSON**

**By**

**Noel Boyle**

**A DISSERTATION**

**Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of**

**DOCTOR OF PHILOSOPHY**

**Philosophy**

**2008**



## **Abstract**

### **PHYSICALISM AND QUALITATIVE FACTS: A CRITIQUE OF FRANK JACKSON**

**By**

**Noel Boyle**

**Frank Jackson believes physicalism is only true if the physical facts about our world entail all of the other facts about our world. Pre-1998 Jackson also believed that certain qualitative facts about our world are not entailed by the physical facts. Post-1998 Jackson believes that physicalism is true and that there are no such qualitative facts. I disagree with Jackson about the sense in which physicalism purports to be a complete metaphysical system, and therefore reject both his pre-1998 and post-1998 views.**

**Pre-1998, Jackson defends his knowledge argument, which features the fictional super-scientist Mary who has learned all the physical facts about human chromatic vision while imprisoned in a black and white room. Upon her release, Mary has her first chromatic experiences. Jackson says it is just obvious that she thereby learns new facts about the world. Thus, there are facts over and above all of the physical facts and physicalism is false in that it provides an incomplete picture of our world.**

**In response, I argue that there is no sense of ‘all the physical facts’ in which it is simultaneously possible to know all the physical facts achromatically, and also true that physicalism asserts that there are no other facts. Specifically, I argue that there is a hidden modal assumption in the story about Mary. By stipulating that Mary knows all the physical facts in a black and white room, Jackson asserts that it is possible to know all the facts that can be countenanced by a physicalist ontology through purely achromatic means. I deny this modal intuition. Turning to Jackson’s own account of the sense in**

which physicalism is committed to the completeness of the physical facts, I argue that the qualitative facts about our world would also be true in a minimal physical duplicate of our world (a physical duplicate, with nothing added).

Contrary to philosophers such as Colin McGinn and Joseph Levine, I argue that models of explanation currently accepted in mainstream philosophy of science can embrace the intuition that certain facts can only be known subjectively. Specifically, Wesley Salmon's causal-mechanical model, Philip Kitcher's unificationist model, and Patricia Churchland's new-wave reductionist model each indicate the direction for a science of qualitative properties.

Jackson would respond to my analysis of the knowledge argument by arguing that physicalism is true if, and only if, all of the true propositions about our world are (in principle) logically entailed by the complete set of true physical propositions. I argue that, as in the knowledge argument, Jackson conflates metaphysical and epistemology issues, and that his persistent refusal to clarify the meaning of 'physical' leads to question begging regarding the physicalist credentials of qualitative facts. I also argue that Jackson fails to show that there is a difference between qualitative facts and biological facts, such that biological facts are placed in a physicalist ontology but qualitative facts are not.

The conclusion and appendix speculate regarding the place of qualitative properties in a physicalist worldview. The conclusion argues that qualitative properties are best understood as finely-grained properties that emerge from a holistic base. The appendix offers a speculative intertheoretic explanatory model connecting European phenomenology, clinical psychology, and neurobiology.

## Dedication

To my hemidecorticated son Ciaran,  
whose hospital bed I sat by when I decided  
to write a dissertation in the philosophy of mind;  
for teaching me to pay attention to the brain.

## Acknowledgments

Jennifer Susse is a wonderful person and a gifted teacher; it is my great fortune to have her as an advisor. Jennifer is the anti-thesis of every clichéd horror story about dissertation advisors. She has been generously supportive in the development of my ideas. She has always been both compassionate and understanding about my struggles simultaneously being graduate student, teacher, and parent. All in all, she has a talent for demanding high academic standards while never being intimidating. She maintained an encouraging and friendly tone even when explaining that various drafts of chapters were “sloppy and windy,” “wrong and misleading,” and even “boring”. This dissertation is better than it would have been had I worked with any other advisor.

Through and through, Michigan State has been a terrific place to study philosophy. I would like to acknowledge three faculty members in particular. At new student orientation Debra Nails’ pitch for her Plato class was inspiring, as was the class itself. Debra’s own contributions to scholarship are enormous. She is also a deeply caring person and critical teacher, one from whom compliments have meaning because she only gives them when they have been well earned. I thank her also for giving me extensive and sound professional advice. I would also like to acknowledge Marilyn Frye. As teacher of the Proseminar, and as then director of the graduate program, she helped me to establish confidence that I could succeed despite my hectic schedule. It still saddens me that no one registered for the social and political philosophy class where she was to be my teaching advisor. It was also a privilege to take a seminar on intelligent design creationism with Rob Pennock. I decided to become a philosopher because philosophy

matters, and Rob is out there “in the trenches” defending the philosophical integrity of public school science education. I also thank Rob for encouraging me to go ahead with plans to teach Freshman at Lyman Briggs from a reading list including Popper, Kuhn, Hempel, Kitcher, Salmon, E. Nagel, and Patricia Churchland; it was fun.

Thanks to the whole of my committee as well: Jennifer, Rob, Fred Gifford, and Barbara Abbot. Having put in so much time writing this dissertation, it is a relief to know that at least four people *have* to read it and tell me what they think.

Most of all, I am inexpressibly grateful to my wife Jessica and my sons, Teagin and Ciaran. Jessica has done as much as I have to make it possible for me to earn a Ph.D. She has tolerated my grouchy and stress-filled disposition, made it possible for me to work when I needed to, and complained very little as I turned our family life to chaos for five years. She is the constant source of reminders about why I chose to become a philosopher, and challenges me to live up to my own ideals. During those inevitable moments of self doubt, it was Jessica who held me, told me that I was brilliant, and convinced me that I shouldn't give up. She also makes sure that I live as balanced of a life as possible, sagely indicating when it was time to put the books away and enjoy time with my friends and children. My son Teagin is simply the most joyous, loving, and hilarious six year old boy that there is. I promised not to tease him for his generous offers to play video games so that I could work on my dissertation, but I could not resist.

Finally, I would like to thank my parents for responding to my intention to pursue a career in philosophy by saying, “great, what the hell is philosophy?” Though not the usual parental response, it is indicative of their unwavering support of all my intellectual endeavors. Indeed, they actively supported my studies long before I did.

## Preface

### Epigraph on epistemology:

“Although I have acquired a thorough theoretical knowledge of the physics of colours and the physiology of the colour receptor mechanisms, nothing of this can help me to understand the true nature of colours.”

--Knut Nordby, an actual colorblind neuropsychologist specializing in chromatic vision. Quoted in Oliver Sacks, *An Anthropologist on Mars*, 36

### Epigraph on metaphysics:

“Tell me one last thing,” said Harry. “Is this real? Or has this been happening inside my head?”

Dumbledore beamed at him, “of course it is happening inside your head, Harry, but why on earth should that mean that it is not real?”

--J.K. Rowling, *Harry Potter and the Deathly Hallows*, 723.

## Table of Contents

Introduction: Epistemology and Metaphysics in Frank Jackson's Work.....	1
I. Jackson's Knowledge Argument.....	1
II. Chapters One through Three.....	9
III. The Entailment Thesis: Jackson's Current Position.....	23
IV. Chapter Four.....	28
 Chapter One: What Happens When Mary Leaves the Black and White Room?.....	33
I. The Second Premise is False.....	34
II. Jackson Equivocation on 'Knows' (or 'Physical', or 'Complete').....	40
A. The Ability-Acquaintance Hypothesis.....	41
B. The Differently Physical Facts Hypothesis.....	48
III. Same Fact, New Mode of Access.....	56
 Chapter Two: Zombie University.....	65
I. Two Parts of the Knowledge Argument.....	65
II. Jackson's Two Intuitions.....	68
III. Rejecting Jackson's Modal Intuition.....	84
 Chapter Three: Qualitative Facts and Scientific Explanation.....	98
I. Accepting the Epistemological Part of the Knowledge Argument.....	98
II. Jackson's Epistemological Intuition and the Explanatory Gap.....	103
III. Knowing <i>That</i> : First Person Epistemic Access and Demarcation.....	105
IV. Knowing <i>Why</i> : In Search of an Explanatory Model.....	110

A. Final Theory Reduction.....	112
B. Nagel-Reduction.....	114
C. New-Wave Reduction.....	117
D. Unificationist Reduction.....	120
E. Causal-Mechanical Explanation.....	124
V. The Knowledge Argument and the Explanatory Gap.....	129
 Chapter Four: Entailment and the Placement of Qualitative Facts.....	135
I. The Entailment Thesis.....	135
A. Linguistic Entailment and the Placement Problem.....	136
B. The Two <i>A Priori</i> Aspects of Entailment.....	140
C. Jackson's Knowledge Argument Clarified, Again.....	147
II. The Entailment Thesis, Qualitative Facts, and the Placement Problem.....	150
A. The Entailment Thesis Does Not Address Earlier Critiques.....	151
B. The Entailment Thesis Fails to Drive a Wedge.....	162
 Conclusion: Solving the Placement Problem	
I. A Speculative Physicalist Metaphysic of Qualitative Properties.....	176
II. A Speculative Physicalist Epistemology of Qualitative Properties.....	184
 Appendix: Neurobiology and Phenomenology.....	191
Reprint from <i>Journal of Consciousness Studies</i> , 15, no.3 (2008): 34-58.	
 Bibliography.....	226



## Introduction:

### Epistemology and Metaphysics in Frank Jackson's Work

Though this dissertation focuses on the work of Frank Jackson my concern is ultimately with a more general philosophical question: what must philosophers who accept physicalism (the view that the entire world is physical) say about the epistemological strategies appropriate for understanding conscious, qualitative experience? Jackson's philosophical career is unusual in that he has changed his mind about central questions (specifically, about whether physicalism is true). Nevertheless, a consistent theme of his work is that physicalists are committed to an *a priori* reductive epistemology. According to Jackson (and David Chalmers), philosophers of mind have two fundamental options available: reductive physicalism or some version of property dualism. Consistently arguing that there is such a dilemma, Jackson has only changed his mind regarding which option is preferable. As a non-reductive physicalist, I believe that Jackson offers a false dichotomy. The falseness of the dichotomy between reductive physicalism and property dualism can be seen most clearly by exploring the relationship between metaphysical and epistemological claims in Jackson's knowledge argument.

#### I. Jackson's Knowledge Argument

The knowledge argument will be referred to so many times that it is invaluable to survey the various accounts of this argument that Jackson has presented in order to provide a formulation of the knowledge argument that both captures Jackson's intention and is a suitable reference point for subsequent discussion.

Jackson opens the initial 1982 article, in which he presented the knowledge argument, by saying that fellow qualia freaks don't give themselves enough credit when admitting that their anti-physicalist position is based on unargued intuition; he thinks that the knowledge argument is a perfectly good anti-physicalist argument that rests on an intuition that is widely accepted.<sup>1</sup> Thus, a central claim of the article is the contrast between the knowledge argument and what he calls the modal argument. The modal argument rests on the disputable (indeed, disputed) intuition that a creature (called a zombie in current jargon) that is a physical and functional duplicate of an actual person but is entirely lacking in conscious experience is possible. In contrast, Jackson's knowledge argument uses thought experiments about Fred and Mary to "present an argument whose premises are obvious to all, or at least to as many as possible".<sup>2</sup> The thought experiment about Mary was almost an afterthought in the original article but has received nearly all the subsequent critical attention. I will briefly describe the Fred thought experiment before quoting the entirety of Jackson's original comments on Mary.

Fred is able to make color discriminations that you and I are unable to make. If you show Fred a certain batch of red apples that normal color discriminators say are all of the same shade, Fred can consistently and easily sort them into red<sub>1</sub> and red<sub>2</sub>. In an attempt to understand what it is like for Fred to discriminate between red<sub>1</sub> and red<sub>2</sub> researchers "acquire all the physical information we could desire about his body and brain, and indeed everything that has ever featured in physicalist accounts of mind and consciousness".<sup>3</sup> Later, researchers discover how to surgically "transplant his optical system into some else".<sup>4</sup> According to Jackson, only after surgery would people know what it is like to for Fred to discriminate between red<sub>1</sub> and red<sub>2</sub>. Since researchers knew

all the physical information *before* surgery, and learned what Fred's conscious experience was like only *after* surgery, there is more to know than all the physical information and, hence, "physicalism is incomplete".<sup>5</sup> In short, the thought experiment about Fred is intended to demonstrate that there is information above and beyond physical information. There is *something* that it is like for Fred to see red<sub>1</sub> that is different than what it is like for Fred to see red<sub>2</sub>. But you can not know what that *something* is on the basis of physical information alone. Thus, there is more information than physical information and physicalism is false in the sense that it is incomplete.

Jackson introduces Mary in order to show that the same argument can be constructed with reference to a "normal" person. The following constitutes very nearly the entirety of Jackson's original comments about Mary.

Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black-and-white room via a black-and-white television monitor. She specialises in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and uses terms like 'red', 'blue', and so on....(It can hardly be denied that it is in principle possible to obtain all this physical information from black-and-white television, otherwise the Open University would of *necessity* need to use color television). What will happen when Mary is released from her black-and-white room or is given a color television monitor? Will she *learn* anything or not? It seems just obvious that she will learn something about

the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had *all* the physical information. *Ergo* there is more to have than that, and physicalism is false.<sup>6</sup>

Clearly, the thought experiments about Fred and Mary each express the same argument: physicalism can not account for what it is like to have a certain qualitative experience and is thus incomplete. From the original 1982 article, I distill the following formulation of the knowledge argument.

1982 Formulation:

1. Before her release, Mary acquires all the physical information that there is to be had. (*ex hypothesi*)
2. Mary learns something when she leaves her black and white room. (“just obvious”)
3. Physical information is incomplete. (from 1 and 2)
4. Therefore, physicalism is false.

Notice that, although the above argument accurately renders Jackson’s comments, it is not, strictly speaking, valid. The conclusion contains the term ‘physicalism’, a term that is not included in any of the preceding premises. To be fair, it should be acknowledged that there is an enthematic premise: that physicalism asserts that there is no information other than physical information. Nevertheless, Jackson passes without comment from ‘physicalism is incomplete’ to ‘physicalism is false’. What the 1982 account of the

knowledge argument lacks is an account of the sense in which physicalism claims to offer a complete metaphysical picture.

In 1986 Jackson responded to some early criticisms of the argument and clarified what Mary exactly did not know inside her black and white room.<sup>7</sup> He still did not describe the sense in which physicalism is committed to completeness, however. Most responses to the 1982 article argued that Jackson had equivocated on the word ‘knowledge’ and not that he was mistaken about the nature of physicalism. Two of his clarifications will be essential to providing a suitable revised formulation of his argument.

First, Mary knew all of the physical “facts”. Physicalism is committed to the proposition that there are no *facts* above and beyond the physical *facts*. The point of the thought experiment about Mary is not that there are limits on Mary’s imaginative capacities, but that there are holes in Mary’s factual knowledge before she leaves the black and white room. I take it that the change from ‘physical knowledge’ to ‘physical facts’ was meant by Jackson as a clarification of what he meant and not a change in his position. By ‘fact’ he means a true proposition about the world; ‘fact’ is the word he uses to designate truth statements about the world.

Second, “the knowledge that Mary lacked which is of particular point for the knowledge argument against physicalism is “*knowledge about the experiences of others*, not about her own [experiences]”.<sup>8</sup> It is no challenge to physicalism to point out Mary has a new experience when she leaves the black and white room<sup>9</sup> and therefore acquires new facts about her own experience. “The trouble for physicalism is that, after Mary sees her first ripe tomato, she will realize how impoverished her conception of the mental life of

*others has been all along*".<sup>10</sup> Inside the black and white room she knew all the physical facts about other people's color experiences, but there was always "a feature conspicuous to them [that was] until now hidden from her".<sup>11</sup>

After making these clarifications, Jackson offers the following "convenient and accurate way of displaying the argument".

Jackson's 1986 Formulation:

1. Mary (before her release) knows everything physical there is to know about other people.
2. Mary (before her release) does not know everything there is to know about other people (because she *learns* something about them on her release).
3. Therefore, there are truths about other people (and herself) which escape the physicalist story.<sup>12</sup>

There are several reasons that the above formulation is not adequate for capturing Jackson's intention. First, the phrases 'knows everything physical' and 'does not know everything physical' are vaguer than is warranted. Jackson has already clarified that the argument pertains to 'physical facts' and replacing 'knows everything physical' with 'knows all the physical facts' is both clearer, and more specific. Second, the conclusion makes reference to 'the physicalist story' and, as this term is not present in either of the premises, the conclusion does not – strictly speaking – follow from the premises. The conclusion that the premises actually support is "there are facts other than all the physical facts"; or, shorter, "there are non-physical facts". Refining Jackson's convenient and

accurate display of argument with the clarifications that Jackson offers yields the following formulation of Jackson's argument.

**Clarified 1986 Formulation:**

1. Mary (before her release) knows all the physical facts there are to know about other people.
2. Mary (before her release) does not know all the facts there are to know about other people (because she *learns* something about them on her release).
3. Therefore, there are facts other than all the physical facts.

One striking feature of this formulation is that it does not conclude that physicalism is false; in fact there is no reference to physicalism in the argument at all. Thus, though the 1986 formulation provides important clarifications of the 1982 formulation, the sense in which physicalism claims completeness is left entirely unarticulated.

In 2004, though he had abandoned the knowledge argument in favor of *a priori* physicalism, Jackson offered the following summary of the knowledge argument as he once defended it.

**2004 Formulation:**

1. Complete physical knowledge is not complete knowledge *tout court*.
2. If physicalism is true then complete physical knowledge is complete knowledge *tout court*.

### 3. Therefore, physicalism is false.<sup>13</sup>

The 2004 formulation is the first I have considered that validly derives the conclusion that physicalism is false. The 2004 formulation is therefore the first to clearly indicate that, on Jackson's view, there are certain things that physicalists must believe that the thought experiment about Mary demonstrates to be false. Premise one of the 2004 formulation is clearly the premise that Jackson intends to establish by means of the thought experiment about Mary. Thus, the 1986 formulations (themselves essentially clarifications of the 1982 formulation) can be seen as the argumentative support for the first premise of the 2004 formulation. Though neither the 1986 nor 2004 formulation provides any argumentative support for the second premise of the 2004 formulation, the overall structure of the knowledge argument is clear enough to present what is for the purposes of this dissertation the official formulation of Jackson's knowledge argument. We will get back to Jackson's account of the sense in which physicalism claims that the physical facts are incomplete.

#### Dissertation Formulation of the Knowledge Argument:

1. Inside the black and white room, Mary knows all the physical facts (about other people).
2. After her release from the black and white room, Mary learns new facts (about other people).
3. Thus, there are facts other than all the physical facts.
4. If physicalism is true, then there are no facts other than all the physical facts.



## 5. Therefore, physicalism is false.

Premise one is established *ex hypothesi*; it is merely stipulative. Jackson considers premise two to be “just obvious”. He acknowledges that the second premise is merely intuitive, but considers it to be a strong and widely held intuition. Line three is a lemma that follows from the first two premises; if Mary already knew all of the physical facts and then learned a new fact then there must be facts other than physical facts. Premise four follows, according to Jackson, from the definition of physicalism. After all, physicalism claims to be a complete metaphysical system. The conclusion follows from premise three and four, by *modus tollens*. Hopefully, I have already shown that the dissertation formulation is both acceptable to defenders of the knowledge argument and true to Jackson’s intentions. I claim that it is true to the original presentation of the thought experiment about Mary, it integrates Jackson’s 1986 clarifications about what Mary didn’t know, and it coheres with Jackson’s most recent comments about the overall structure of the argument he once defended. It is also clearly valid. For the sake of convenience and clarity all subsequent use of ‘knowledge argument’ in the dissertation refers to this formulation of the argument.

## II. Chapters One through Three

The first three chapters of the dissertation explicitly concern the knowledge argument. In Chapter One I categorize and critically survey the main lines of physicalist responses to the knowledge argument. Chapter Two is my account of why the knowledge argument fails to establish the metaphysical conclusion that physicalism is false. Chapter

Three is my account of why the knowledge argument fails to establish the epistemological conclusion that qualitative states can not be explained by the natural sciences. Chapter Four is best explained a little later.

### *Chapter One*

Physicalist response to the knowledge argument has focused on the second premise; there has been detailed discussion of precisely what happens to Mary when she leaves the black and white room. Some (Paul Churchland, Daniel Dennett) say she learns nothing; others (David Lewis, Earl Connee) say she learns something non-factual, others (Loar) say she gains a new conceptual understanding of the same properties, and yet others (Terence Horgan, Torin Alter) say she learns new physical facts. I argue that, for various reasons, none of these responses decisively answer Jackson's knowledge argument. The consistent theme that weakens these responses, if only for strategic reasons, is that they are all willing to grant that Mary possesses complete physical knowledge in some meaningful sense of physical knowledge.

Once it is granted that a physically omniscient achromatic knower is possible (in at least some sense of 'physical') Jackson's persistently ambiguous use of 'physical' leaves him in a rhetorically strong position. Nearly all the responses to the thought experiment accuse Jackson of some kind of equivocation. The most common strategy is to argue that Jackson employs one sense of 'knowledge' in the first premise and a different sense of 'knowledge' in the second premise. Since the first two premises do not share a term, the sub-conclusion (line three) does not follow. Given Jackson's lack of clarity about the meaning of 'physical' he can easily deny that the distinction upon which

the equivocation charge is made accurately reflects his intention. That is, Jackson criticizes the physicalists' specification of the knowledge that Mary has inside the black and white room instead of providing his own specified account. Put differently, physicalists have tended to respond to the knowledge argument by distinguishing, in various ways, between "all physical knowledge<sub>1</sub>" and "all physical knowledge<sub>2</sub>". By remaining ambiguous about what he means by 'all physical knowledge' Jackson can claim that Mary knows more than all physical knowledge<sub>1</sub> (as in his response to Horgan) or that all physical knowledge<sub>1</sub> implies all physical knowledge<sub>2</sub> (as in his response to Loar) or that all physical knowledge<sub>2</sub> just isn't the post-release event that he had in mind (as in his response to Lewis). In the end, physicalists' attempts to fill in for Jackson's ambiguity backfire.

## *Chapter Two*

Instead of focusing on the second premise, physicalists ought to focus on the first premise. Considering the sense in which physicalism is committed to the completeness of physical knowledge, physicalists can (and I think should) respond that it is not possible to know all the physical facts in a black and white room. I don't mean the mundane observation that practical barriers would impede Mary's progress, but that there are facts about the physical world that could never be known (even by a God) except by means of chromatic experience.

Uncontroversially, Jackson needs to establish both an epistemological and a metaphysical claim. As the knowledge argument is usually understood, the first two premises (that is, the thought experiment about Mary) establish the epistemological claim

and the fourth premise establishes the metaphysical claim. However, seeing the argument in this way overlooks the connection between the first and fourth premises of the argument. By way of reminder, the two premises at issue are:

1. Inside the black and white room, Mary knows *all the physical facts*.<sup>14</sup>
4. If physicalism is true then there are no facts other than *all the physical facts*.

Both premises include the phrase ‘all the physical facts’, and the knowledge argument is valid only if ‘all the physical facts’ means the same thing in the two lines. Put somewhat differently, it is not enough that Mary know all the physical facts in any old sense of ‘all the physical facts’. For the knowledge argument to be valid, Mary must know all the facts *as far as physicalism is concerned; all the facts countenanced by physicalist ontology*.

When the crucial link between the first and fourth premises is highlighted, it becomes clear that the thought experiment about Mary is not purely epistemological: the issue of Mary’s stipulated pre-release knowledge can not be separated from questions about the meaning of ‘physical’ and, thereby, about the standards for including facts in a physicalist ontology. The purely epistemological intuition operating in the argument is that there are certain facts about the world that can only be known by means of chromatic experience. Whether there are facts that can only be known through qualitative experience is one issue; whether such facts are consistent with physicalist ontology is quite another. Put differently, Jackson’s epistemological intuition can be fully expressed by imagining Harry, Mary’s neighbor. Locked in a black and white room identical to Mary’s for his entire life, Harry is otherwise just an ordinary bloke who knows no more

than any other average person. Making no reference to ‘physical’ or ‘physicalism’, it is possible to fully express Jackson’s epistemological intuition by saying that there are things that Harry just couldn’t know. It is entirely an open question whether the intractable lapses in Harry’s knowledge represent a threat to physicalism – an open question that can only be addressed by analysis of physicalism’s epistemological commitments.

In short, Jackson needs a reason for claiming that qualitative facts are not physical that is independent of the thought experiment. To suggest, as he does, that we know that qualitative facts are not physical because, *ex hypothesi*, Mary knew all the physical facts while still locked in the black and white room begs the question about the physicalist credentials of chromatically accessed facts. Of course, Jackson can stipulate whatever he wants to about Mary, but to suggest that she knows all the facts countenanced by physicalist ontology inside a black and white room is to assert that it is possible to know all the physical facts achromatically. Thus, the first premise of the argument hides the following modal intuition: omniscience regarding the physical is possible achromatically (and, by extension, is possible without qualitative experience).

Of course, Jackson might defend this modal claim by defining ‘physical’ in such a way that excludes qualitative facts. Yet, he persistently refuses to define ‘physical’. Explicitly stating that he is not offering a definition of ‘all the physical facts’ Jackson indicates that he means something like a completed physics, chemistry, biology, and all that follows from it.<sup>15</sup> I argue that Jackson’s characterization of ‘all the physical facts’ in terms of a completed (and thus idealized) science is flawed for two reasons. First, claims that *x* could not be known by an idealized science are automatically suspect; it is

generally unwise to bet against the explanatory power of future, let alone idealized, science. I see no reason to believe that chromatically accessed facts are beyond the explanatory reach of idealized science. Second, physicalism is not committed to the belief that all facts are scientifically knowable. Science is the deeply human practice of seeking an objectively true understanding of the natural world, and does not assert dominion over all truths about the physical world. There is nothing inherently anti-physicalist in the claim that some facts about the world are beyond the reach of science.

Since Jackson inadequately defines ‘all the physical facts’ when describing what Mary knows inside the black and white room it is necessary to turn to his more recent writings, in which he explicitly provides an account of physicalism’s claim to completeness. Jackson asserts that physicalism is committed to the following supervenience thesis: “a minimal physical duplicate of our world is a duplicate of our world *simpliciter*”.<sup>16</sup> A physical duplicate of our world is what you would get if you were to build a world that contains all of the physical property instances and relations of the actual world; the use of ‘minimal’ means simply that nothing else should be added. Physicalism is committed to the claim that a perfect physical duplicate of the actual world, with nothing at all added to it, is in no way different from our actual world. Though I disagree with Jackson about the implications of the minimal supervenience thesis (see summary of Chapter Four below), I agree that the thesis captures a genuine commitment of physicalism. Interpreting ‘all the physical facts’ as ‘all the facts that also obtain in a minimal physical duplicate world’ is an excellent way of expressing the sense in which premise four of the knowledge argument is true. The question is then whether the stipulation of the first premise (Mary knows all the physical facts achromatically) is a

stipulation of something that is possible when ‘all the physical facts’ is interpreted as ‘all the facts that also obtain in a minimal physical duplicate world’. That is, the following modal claim (stated more precisely than above) is implied by the first premise of the knowledge argument: it is possible to know all the facts about a minimal physical duplicate world without chromatic experience. As a physically omniscient achromatic knower, Mary is only possible if there are no facts about a minimal physical duplicate world that can only be known on the basis of chromatic experience.

I endorse a narrow reading of Jackson’s epistemological interpretation: there are facts about the qualitative nature of experience that can only be known chromatically (I refer to such facts as ‘qualitative facts’). However, this epistemological intuition will lead to the conclusion that physicalism is false only Jackson’s modal claim is also true. The knowledge argument is generally understood in terms of an epistemological thought experiment and a claim about the nature of physicalism; it is better to see the knowledge argument in terms of an epistemological claim (there are qualitative facts) and a modal claim (it is possible, by achromatic means, to know all the facts about a minimal physical duplicate world). I accept the epistemological claim and reject the modal claim.

My strategy for rejecting Jackson’s modal claim begins by highlighting its similarity to Chalmers’ modal intuition that zombie worlds are logically possible. As Chalmers uses the term, a zombie world is a perfect physical and functional replica of our world; were you to visit a zombie world you would notice no difference between that world and ours. However, the creatures who live there (zombie twins of all the people in the actual world) have no qualitative experiences; by definition, there is nothing that it is like to be a resident of a zombie world. As I see it, Jackson’s modal claim and

epistemological claims are simultaneously true if, and only if, a minimal physical duplicate world is a zombie world. If a minimal physical duplicate of our world is a zombie world, then the knowledge argument does establish the falsity of physicalism: inside the black and white room Mary learns everything there is to know about a minimal physical duplicate world and, when released, learns the qualitative facts about our world. If a minimal physical duplicate world is not a zombie world, then the knowledge argument does not establish the falsity of physicalism: to learn everything about a minimal physical duplicate world is also to learn about the qualitative facts of that world and is thus to have the qualitative (specifically, chromatic) experiences that are necessary for learning qualitative facts. In which case the first premise does not assert a genuine possibility. But the claim that a physical and functional replica of our world, with nothing else added, lacks qualitative facts *just is* Chalmers' zombie intuition. To assert that zombies are possible is to claim that qualitative facts are not among the physical and functional facts about our world.

In one sense, making the case that Jackson's knowledge argument relies on an intuition similar to Chalmers' intuition that a zombie world is possible is enough to defeat Jackson's argument; the purpose of Jackson's original article presenting the knowledge argument was to provide an anti-physicalist argument that does not rely on a disputed modal intuition.<sup>17</sup> In a more complete sense, defeating the knowledge argument requires a refutation of Jackson's modal intuition that one can learn everything about a minimal physical duplicate world through completely achromatic means. As Chalmers points out, arguing about modal intuitions usually amounts to nothing more than elaborate question begging.<sup>18</sup> As Chalmers sees it, a zombie world *seems* logically possible; one who denies



the possibility of a zombie world must point out the sense in which the description of a zombie world involves an implicit contradiction or a misdescription. In the last part of Chapter Two, I argue that Jackson's modal intuition indeed entails a contradiction. 'Contradiction' might seem too strong here, but it is important to note that Chalmers' intuition is that zombies are logically, as opposed to naturally, possible. Asserting that something is naturally possible is to assert that it could actually happen in our world, asserting logical possibility is merely to assert that it can be imagined without contradiction.<sup>19</sup> I accept Chalmers' argument that physicalism is committed to a logical supervenience thesis and I am prepared to defend the claim that premise one involves a contradiction. In fact, the contradiction can be stated quite simply: Jackson explicitly states that Mary knows all the physical facts, and (because she is locked in a black and white room) Jackson implicitly states that Mary does not know all the physical facts (because Jackson's modal claim is false). A replica of me in a minimal physical duplicate world is (like me) currently sitting outside in the sunshine, looking at the green grass, and drinking a pint of fine ale. A perfect physical replica of me has a brain that is in precisely the same state as my brain: the warm sunshine, green grass, and fine ale are having precisely the same neural effects on him that they are having on me. I see no grounds whatsoever to deny that such a perfect physical replica of me could fail to also be a qualitative replica of me. What it is like for me to taste this fine ale is exactly the same as what it is like for my minimal physical replica to taste it because the physical states that make true my qualitative states will make true the same qualitative states in him. Lucky for him, he is not a zombie.

So, if Mary were to know all of the facts about my minimal physical replica, she would have to know what it is like for my replica to be in various qualitative states, including chromatic states. If Mary is locked in a black and white room, she won't be able to know what it is like for my minimal physical replica to (for instance) see red. Thus, the assertion that she knows all the physical facts about a minimal physical duplicate world while locked in a black and white room is false in that it asserts a contradiction.

Such arguments will fail to convince defenders of the zombie intuition. However, physicalists routinely assert the intuition that a physical replica of me can not fail to also be a qualitative replica and should thereby reject Jackson's modal intuition.<sup>20</sup> By denying Jackson's modal claim, physicalists can acknowledge that the knowledge argument advances an interesting and worthwhile epistemological intuition in a way that is no threat to physicalism.

### *Chapter Three*

It is important to distinguish my response to the knowledge argument from Colin McGinn's or Joseph Levine's response. I agree with McGinn and Levine that the knowledge argument establishes an important epistemological point but fails to establish the falsity of physicalism. However, McGinn and Levine argue for a broader interpretation of the knowledge argument's epistemological claim. According to McGinn, the thought experiment about Mary is an effective way to show that no amount of third-person, scientific information can provide insight into the kind of first-person perspective essential for understanding qualitative states.<sup>21</sup> It is McGinn's view that qualitative states are complex physical states that will forever remain mysterious. According to Levine, the

thought experiment about Mary establishes the existence of an “explanatory gap”; physicalist theories of mind are deeply flawed in that consciousness escapes their “explanatory net”.<sup>22</sup> It is Levine’s view that qualitative states are physical but can not be explained by the physical sciences.

In order to examine views, such as those advanced by McGinn and Levine, and to avoid the question begging created by denying Mary chromatic experience, I recast the knowledge argument in purely epistemological terms. Imagine Mary’s twin sister Hermy, who is free to explore the world by whatever means she finds most suitable and is given the task of writing a dissertation titled, “The Correct Scientific Explanation of What it is Like to See Red”. She is told by her advisors that she can not accept an eliminativist approach to consciousness; she must accept the intuition that there are genuine qualitative facts that can only be known by means of qualitative experience. If Levine and McGinn are correct, then Hermy is doomed to be the stereotypical graduate student who chooses an overly ambitious doctoral thesis and stays in graduate school forever. Paralleling McGinn’s and Levine’s claims, there are two approaches one might take in arguing that Hermy will never be able to graduate. First, one might argue that, because qualitative facts can only be accessed from the first-person, and scientific facts are accessed from the third-person, there can never be a science of qualitative facts. Second, one might argue that it will be impossible for her to establish scientifically credible inter-theoretic connections between phenomenal psychology and lower levels of scientific discourse.

The claim that no amount of subjective first-person knowledge can meet the scientific standard of objective third-person knowledge is clearly an issue of demarcation. That is, McGinn’s claim relates to distinguishing scientific reasoning from non-scientific

reasoning. McGinn's view is that the first-person nature of qualitative facts precludes phenomenal psychology from being a genuine science. Admittedly, the kind of knowledge required to provide an accurate account of qualitative facts does not live up to the scientific ideal of directly observable data. But neither does scientific knowledge about the ancestral connection between horses and zebras, or the nature of black holes. Qualitative facts are inherently subjective, but that does not prevent knowledge of them from being objective in the sense relevant to demarcation. Hermy will likely need to have colored patches sprinkled throughout her dissertation (which themselves can be properly understood only by someone with the capacity for chromatic vision), but there is no reason to suspect that Hermy's research will necessarily be tainted by her personal biases any more than any other scientist's work. She will, for instance, precisely describe in her dissertation the extent to which your experiences will resemble hers when you look at the color patches in her dissertation. Hermy's work would even be falsifiable; though the central propositions delineating qualitative facts will require a first-person perspective, they can be falsified by anyone willing to check them against their own first-person experiences or other people's accounts of their own first-person experiences.

To me it seems odd to say that Hermy's work could never be scientific because of the first-person perspective that is required to understand qualitative facts. All scientific facts are subjective in that sense. You can never know what it's like to see red unless you have seen red; but you can never know what it's like to understand some fact of chemistry unless you have understood it for yourself. Knowledge is inherently in the first-person perspective.

Regarding Levine's position that the knowledge argument establishes the incompleteness of physicalist epistemology, I argue that it depends on which model of scientific explanation Hermy is told to use. If Hermy uses an ontologically simplifying style of explanation in which there is ultimately only physics, such as that advocated by Steven Weinberg, then indeed a variation of the knowledge argument can establish the failure of science to explain qualitative facts. However, Weinberg has few fans in the philosophy of science community because precious little actual science lives fits the model he envisions. If Hermy is told to provide an Ernest Nagel-style reduction, based on the deductive-nomological model of explanation, there is some reason for optimism. Nagel recognizes that bridge laws, which connect the vocabulary of different levels of scientific analysis, are not *always* ontologically simplifying. Hermy would have to discover the "conditions for the occurrence" of qualitative facts – and do so in the language of physics, chemistry, and biology.<sup>23</sup> I admit there is a strong intuition that Hermy would not finish her dissertation if she is compelled to provide Nagel-style reductions. However, the deductive-nomological model on which Nagel's view of reduction relies is no longer accepted in mainstream philosophy of science due to the existence of compelling counter-examples. On any model of intertheoretic scientific explanation that is currently viable there is ample reason to be optimistic about a scientific explanation of qualitative facts. I offer three examples of such explanatory models (based on the work of Patricia Churchland, Philip Kitcher, and Wesley Salmon) indicating how each model indicates a direction for Hermy's research. Contrary to popular misunderstandings, Churchland's new-wave reductionist model does not eliminate higher levels of scientific discourse in favor of a purely neuroscientific

explanation. Instead, Churchland advocates a co-evolutionary model of intertheoretic explanation, in which scientific psychology and neuroscience constrain and inform each other as they move towards “reductive consummation”.<sup>24</sup> Hermy should also be optimistic if she is told to use Kitcher’s unificationist model of explanation. According to Kitcher, explanation involves showing that the propositions and argument forms used in a particular scientific discipline (phenomenal psychology, in Hermy’s case) fit within the explanatory nexus created by the rest of scientific knowledge.<sup>25</sup> Perhaps Hermy should be most optimistic about finishing if she is told use Salmon’s causal-mechanical model of explanation. According to Salmon, an event is explained when its causes have been delineated.<sup>26</sup> Naturally, the main burden of Salmon’s work is to reliably distinguish between genuine causal processes and pseudo-processes. Though I will here avoid the details regarding how Salmon attempts to meet this burden, there is no particular reason (from either a naïve or Salmonian perspective) to think that the process starting with stimulation of the cones in the eye, for example, and ending with a qualitative color experience is a pseudo-process.

Despite the current prominence of these last three models of explanation in philosophy of science, nearly all of the debate in philosophy of mind still presumes that a scientific explanation must meet the standard of a deductive-nomological model of explanation. Were the debate within philosophy of mind to reflect an up to date understanding of scientific explanation, the problems associated with a science of qualitative facts would not seem so intractable.

Taking Chapters Two and Three together, I suggest that Jackson’s conflation of metaphysical and epistemological issues can account for much of the appeal of the

knowledge argument. When the metaphysical and epistemological questions are appropriately separated, we will see that the knowledge argument provides insufficient reason to believe that either physicalism or science is condemned to incompleteness. When Jackson's epistemological intuition is stated in an appropriately narrow way, the result is more of an inspiration about how to move forward with a physicalist account, and scientific explanation of, consciousness than it is a path to dualism.

### III. The Entailment Thesis: Jackson's Current Position

In 1998, Jackson changed his mind: he rejected the knowledge argument and embraced what he calls *a priori physicalism*.<sup>27</sup> Before summarizing Chapter Four, it will be necessary to explain where Jackson did (and did not) change his mind.

Recall that, as Jackson sees the knowledge argument, there are two crucial claims. First, there is the epistemological intuition that no amount of physical information can tell you what it is like to (for instance) see red. Second, there is the metaphysical claim that physicalism is committed to denying the epistemological intuition. Jackson still believes in the metaphysical claim of the knowledge argument. He now believes, however, that his pre-1998 epistemological intuition rests on an illusion (for a description of the illusion he believes is involved, see Chapter One). In fact, the 2004 formulation of the knowledge argument that I presented earlier can serve as an effective way to demonstrate where Jackson has (and hasn't) changed his mind.

#### 2004 Formulation of the Knowledge Argument

1. Complete physical knowledge is not complete knowledge *tout court*.

2. If physicalism is true then complete physical knowledge is complete knowledge *tout court*.
3. Therefore, physicalism is false.

Pre-1998, Jackson accepted premises one and two of the 2004 formulation and concluded, by *modus tollens*, that physicalism is false. On Jackson's current view, the empirical evidence collected by the cognitive sciences overwhelmingly favors physicalism (unfortunately, he has said almost nothing to explain his support for this claim).<sup>28</sup> Once Jackson combines the truth of physicalism with his original claim about the completeness of physicalism the falsity of Jackson's original epistemological intuition follows by *modus ponens*. The philosophical work then to be done is to explain why the illusory intuition of premise one seems so appealing. Thus, acceptance of physicalism motivates Jackson's rejection of the thought experiment about Mary – not the other way around.

In that Jackson's current position relies on the same vague notion of complete physical knowledge that his former position did, most of the arguments I have made against Jackson's former position apply equally to his new position. What I think is the flaw in Jackson's knowledge argument – that it falsely implies that physicalism is committed to the completeness of the facts established on the basis of the natural sciences – is equally a feature of his new position.

Both in 1982, and twenty-five years later, Jackson would respond to Chapter Two and Chapter Three by arguing that I misunderstand the sense in which physicalism is committed to completeness. Starting a few years before Jackson's conversion to



physicalism, most of Jackson's work defends conceptual analysis. Jackson argues that physicalism is committed to the in principle existence of an *a priori* entailment of any given psychological fact from the complete set of purely physical facts (a position he calls *the entailment thesis*). Thus, Jackson would respond my Chapter Two by arguing that if qualitative facts can only be known by means of qualitative experience then qualitative facts are not *a priori* entailed by the facts of physics, biology, and chemistry. And, therefore, qualitative facts have no place in a physicalist ontology. Jackson would respond to my Chapter Three by arguing that new-wave reductionist, unificationist, and causal-mechanical models of explanations may have heuristic value but do not themselves generate the kind of relations that are necessary to support physicalist ontology.

If the entailment thesis is false, his current position is affected just as much as his former position. And since the part of Jackson's position that he hasn't changed is the part that I disagree with, then (in broad strokes) the critiques I made of Jackson's knowledge argument should also apply to his defense of conceptual analysis. Indeed, just as with the knowledge argument, I argue that Jackson's defense of conceptual analysis (particularly the entailment thesis) conflates metaphysical and epistemological issues, has an excessively narrow view of explanatory reduction, and suffers from Jackson's consistently ambiguous use of 'physical'.

### *The Argument from Conceptual Analysis*

As Jackson points out, physicalists are committed to the kind of completeness that is expressed in the following supervenience thesis: "any world which is a minimal

physical duplicate of our world is a psychological duplicate of our world”.<sup>29</sup> ‘Minimal’ is essentially a stop clause for constructing other possible worlds; it means that any world that is a physical duplicate of ours *and contains nothing else* is a duplicate *simpliciter* of our world. As Jackson puts it, “a minimal physical duplicate of our world is what you would get if you – or God, as it is sometimes put – used the physical nature of our world (including of course its physical laws) as a recipe in this sense for making a world”.<sup>30</sup> The need for ‘minimal’ is uncontroversial in that a possible world which is a physical duplicate of our world and also contains ectoplasm is not a duplicate *simpliciter* of our world. Such a world is no challenge to physicalism.

According to Jackson, if physicalism is true then any possible world which is an exact replica of the actual world, and has nothing else added to it, is an exact replica of this world *in every way*. Purely by holding the physical facts constant across worlds (and adding nothing non-physical) the economic, political, ethical, theological, and psychological facts would also be held constant. His argument that physicalism is committed to this supervenience thesis is straightforward and convincing. First, if this supervenience thesis is false then physicalism is false: there would then be something in our world which is not contained in a minimal physical duplicate of our world; it would follow that there is something non-physical in our world and physicalism would be incomplete. Second, if physicalism is false then the supervenience thesis is false: if physicalism is false then our world contains something non-physical that is not contained in a minimal physical duplicate world, and thus such a world is not a duplicate *simpliciter* of our world. Therefore, physicalism is true if and only if the minimal supervenience thesis is true.

According to Jackson, accepting the minimal supervenience requires an acceptance of the “entailment thesis”: “psychological facts have a place in the physicalists’ world view if and only if they are entailed by some true, purely physical statement”.<sup>31</sup> Put more precisely, for physicalism to be true the sentence containing the complete set of true propositions concerning the physical nature of the world must *make true* any given true proposition regarding the psychological nature of the world.<sup>32</sup> Further, this *making true* of psychological facts by physical facts must be done *without further assistance*; that is, the entailment must be *a priori*. Of course, the ‘complete set of true propositions concerning the physical nature of the world’ can not be actually constructed; Jackson calls it “a sentence in some idealized language constructed from the materials that serve to give the full, complete account of the physical sciences”.<sup>33</sup> Nor does the entailment thesis depend on the ability of any human intellect to distill psychological facts from the complete set of true physical propositions. The entailment thesis is a matter of macroscopic facts (including psychological facts) being determinable *a priori in principle* (by God, as it is sometimes put) from a complete physical story.

#### *Explication of Premise Four of the Knowledge Argument*

The entailment thesis specifies the sense in which physicalism is committed to completeness, and thus serves as a definition of Jackson’s notion of a complete set of facts. In short, because physicalism is committed to the minimal supervenience thesis, all facts about the world must follow *a priori* from those facts that are expressed in terms of the limited number of entities and properties privileged by physicalism. I can now present a more or less formal version of Jackson’s argument for the entailment thesis, such that

the conclusion is the fourth premise of the dissertation formulation of the knowledge argument. By the way, the following argument mixes letters and numbers in order to highlight that the conclusion is the fourth premise of the knowledge argument; it is mere coincidence that it is also the fourth line of this argument.

#### Argument Supporting Premise Four of the Knowledge Argument

- a) If physicalism is true, then the minimal supervenience thesis is true.
- b) If the minimal supervenience thesis is true, then the entailment thesis is true.
- c) If the entailment thesis is true, then there are no facts other than all the physical facts.
- 4. Therefore, if physicalism is true, then there are no facts other than all the physical facts.

#### IV. Chapter Four

As is already clear from my summary of Chapter Two, I agree with Jackson that the minimal supervenience thesis reflects a genuine commitment of physicalism; (a) in the above argument is true. Also, I take it that (c) is uncontroversial. It is (b) that I reject. On my view, the entailment thesis is a sufficient, but not necessary, condition for placing qualitative facts within a physicalist worldview. Furthermore, entailment is a condition that is rarely (if ever) actually met. In Chapter Four I argue that, far from adequately responding to the objections I raised against the knowledge argument in Chapters Two and Three, the entailment thesis provides further examples of Jackson making the same mistakes that he did in the knowledge argument: that is, he conflates metaphysical and

epistemological issues, he begs the question by refusing to define ‘physical’, and he presents an excessively narrow view of scientific explanation.

In Chapter Two, I argued that Jackson’s thought experiment about Mary establishes a sense of epistemological incompleteness and thereby asserts a sense of metaphysical incompleteness. Likewise, as I see it, Jackson’s defense of the entailment thesis accurately establishes the sense in which physicalism is committed to metaphysical completeness, and then, inaccurately in my view, asserts that the sense in which physicalism is committed to epistemological completeness has been established. I agree with a strictly metaphysical reading of the minimal supervenience thesis. That is, I agree with Jackson that physicalists must hold that “any psychological fact about our world is entailed by the physical nature of our world”<sup>34</sup>. I disagree with Jackson that it follows that physicalists must hold that the psychological story about our world must be entailed by the story about our world that emerges from physics, chemistry, and biology. The former is a metaphysical claim about the relationship between the facts about our world (facts that are true whether we know them or not). The latter is an epistemological claim about the relationship between sentences derived from different modes of knowing. Put differently, the former is a claim about metaphysical entailment: the lower level facts make the psychological facts true. The latter is a claim about epistemological entailment: the psychological facts are conclusions of deductive arguments in which the lower facts serve as premises. As I see it, the metaphysical claim simply does not entail the epistemological claim. There is no contradiction involved in simultaneously asserting that 1) once God made true all of the microphysical facts there was no more work for her to do, and 2) we can never, even in principle, say the interesting things there are to be said

about qualitative facts merely by presenting long and complicated conjunctions of microphysical facts. In Chapter Four, I argue that epistemological entailment is a sufficient, but not necessary, reason for accepting metaphysical entailment.

In Chapter Two, I argue that Jackson's refusal to define 'physical' leads to question begging regarding the physicalist credentials of qualitative facts. Even in his most recent papers, Jackson says that he will pass over what is meant by 'physical', asserting that doing so will allow him to avoid the tangential controversy of whether the traditional privileging of physics, chemistry, and biology is justified.<sup>35</sup> But this is not a tangential issue. Jackson claims that psychological facts must be entailed by the complete set of physical facts, but the only grounds on which he asserts that psychological facts are not themselves physical is a result of a traditional prejudice among physicalists.

Whatever biases in favor of physics have tended to dominate analytic philosophy, few psychologists would be willing to accept that theirs is a non-physical discipline. And if psychology is recognized as a physical science in its own right, there is no question of psychological facts needing to be entailed by physical facts.

In Chapter Three, I argue that philosophers such as McGinn and Levine, who use the knowledge argument to establish the incompleteness of physicalist epistemology, have an excessively narrow view of scientific explanation. Jackson commits the same mistake when he claims that psychological facts must be entailed by physical facts. As Ned Block and Robert Stalnaker point out, such entailments are not to be found in reductive explanations currently accepted in fields such as biology.<sup>36</sup> While it is uncontroversial that life can be reductively explained in terms of digestion, respiration, locomotion, and so on, it seems utterly hopeless to suggest that the concept of life can be

entailed by the collection of physical and chemical facts regarding digestion, respiration, and locomotion. Not only are such entailments not currently available, there is no reason to suggest that they ever will or could be. Jackson's entailment thesis sets a standard for explanation of qualitative facts that is higher than the generally accepted standards employed in explanation of life. Again, while entailment from physics, chemistry, and biology is undoubtedly a sufficient condition for placement of higher level facts in a physicalist metaphysical system, there is every reason to deny that it is a necessary condition.

## Conclusion

In the conclusion and, more so, the appendix to the dissertation, I speculatively offer an account of how phenomenological psychology stands in an intertheoretic explanatory relationship with psychology and neurobiology. But speculation must be preceded by critical analysis of the work of philosophers who are far better than myself.

---

## Notes to Introduction

1 Jackson, "Epiphenomenal Qualia," 40.

2 Ibid, 40.

3 Ibid, 42.

4 Ibid, 42

5 Ibid, 42.

6 Ibid, 42-43.

7 Jackson, "What Mary Didn't Know," 52.

8 Ibid, 52.

9 As many have pointed out (see Graham and Horgan, "Mary, Mary, Quite Contrary," 61), there are a number of obvious practical problems to the Mary thought experiment. For example, simply locking her in a black-and-white room will not prevent her from having chromatic experiences of her own. She might still have colored dreams, or color experiences caused by rubbing her eyes too vigorously, her own body would have to be painted black-and-white. And so on. For another example, if she was somehow prevented from ever having color experience, her color processing system would atrophy to such a degree that she would be incapable of color experience after her exhaustive studies. Of course, these practical issues do nothing but distract from the interesting philosophical questions Jackson raises. Though some have suggested pigment removing contact lenses or surgical resection of color processing cortical tissue in order to address the practical problems, it is far simpler to just give Jackson a pass on all these issues and just assume that

---

Mary's captors found *some* way to prevent her from having color experiences without destroying her capacity to have them.

10 Jackson, "What Mary Didn't Know," 52.

11 Ibid, 53.

12 Ibid, 54.

13 Jackson, "Preface," xix.

14 In this, and subsequent, references to the first and second premises of the knowledge argument, I leave out the clause "about other people" for the sake of brevity. After all, the point of Jackson adding the clarification that Mary's knowledge is *about other people* is to highlight that the mere fact that Mary has a new experience upon release is no threat to physicalism. Indeed, Mary is claimed by Jackson to learn new facts about chromatic vision in general upon release – her own and other people's. Jackson highlighted the "about other people" clause to highlight that she *learns new facts*, not merely *has new experiences*. In my formulation of the knowledge argument, Jackson's intention is taken into account in the consistent use of learning 'facts' to refer to post-release epistemic events.

15 Jackson, "What Mary Didn't Know," 51.

16 Jackson, "Armchair Metaphysics," 28.

17 Jackson, "Epiphenomenal Qualia," 43.

18 Chalmers, *The Conscious Mind*, 96.

19 By way of examples, a mile-high unicycle is logically but not naturally possible; a cubical sphere is neither logically nor naturally possible; world peace is, however unlikely, both logically and naturally possible.

20 For a paradigmatic example of physicalist rejection of Chalmers' modal intuitions see Hardcastle, "Why of Consciousness," 11.

21 McGinn, "Can We Solve Mind-Body Problem," 533.

22 Levine, "On Leaving Out What It's Like," 553.

23 Nagel, "Issues in Reduction," 914.

24 Patricia Churchland, *Neurophilosophy*, 283. See also Patricia Churchland, "Do We Propose to Eliminate Consciousness?," 298.

25 Kitcher, "1953 and All That," 368. See also Kitcher, "Explanatory Unification," 433.

26 Salmon, "At-At", 194.

27 Jackson, "The Case for *A Priori* Physicalism".

28 Jackson, "Mind and Illusion," 421.

29 Jackson, "Armchair Metaphysics," 30.

30 Ibid, 28.

31 Ibid, 32.

32 See Jackson, "Armchair Metaphysics," 31; Jackson, *Metaphysics to Ethics*, 25; Jackson and Chalmers, "Conceptual Analysis and Reductive Explanation," 328.

33 Jackson, *Metaphysics to Ethics*, 25.

34 Jackson, "Armchair Metaphysics," 31.

35 Ibid, 26.

36 Block and Stalnaker, "Conceptual Analysis and Explanatory Gap," 29.



## Chapter One:

### What Happens To Mary When She Leaves the Black and White Room?

#### (Physicalist Responses to Mary)

Conventional wisdom in philosophy of mind is that Jackson's knowledge argument hinges crucially on the question of what happens to Mary when she leaves the black and white study. By way of reminder, by 'knowledge argument', I mean the following argument:

#### The Knowledge Argument

1. Inside the black and white room Mary knows all the physical facts.
2. After her release from the black and white room Mary learns new facts.
3. Thus, there are facts other than all the physical facts.
4. If physicalism is true, then there are no facts other than all the physical facts.
5. Therefore, physicalism is false.

Physicalist response to the knowledge argument has, in various ways, focused on the second premise. The boldest response, proffered by Daniel Dennett, Paul Churchland, and post-1998 Jackson, is to deny that Mary learns anything at all when she leaves the black and white study. The most common response is to argue that Jackson equivocates by using one sense of 'knowledge' to refer to what happens inside the black and white room and another sense of 'knowledge' to refer to what happens when Mary leaves the black and white room. Loar's response is that Mary gains a new mode of epistemic

access to the same properties she knew about in her black and white room. In this chapter, I argue that none of these responses definitively answer the knowledge argument because each of them grants that there is a meaningful sense (if not the relevant sense) in which Mary has all the physical facts inside the black and white room. In Chapter Two, I argue that physicalists are better off denying the possibility of a physically omniscient achromatic knower. That is, whereas physicalist responses have tended to grant the first premise of the knowledge argument and focus on the second, I will argue that physicalists ought rather to focus on denying the stipulated first premise.

#### I. The Second Premise is False

The most direct response to the knowledge argument is to deny that Mary learns anything new upon leaving the black and white room. According to Dennett's and Churchland's view, Mary could learn qualitative facts while still inside the black and white room. According to Jackson's post-1998 view, when Mary leaves the black and white room she will have new experiences, but her new experiences are the acquisition of new illusions, not new facts about the world.

*Dennett and Churchland: Mary has no new epistemic event at all after leaving the black and white room.*

Dennett chides other philosophers for “simply not following directions” in that they do not even attempt to seriously consider what “all physical knowledge” would include.<sup>37</sup> As Dennett puts it, critics and defenders of the knowledge argument alike, “just imagine that she knows lots and lots – perhaps they imagine that she knows everything

that anyone knows *today* about the neurophysiology of color vision. But that's just a drop in the bucket, and it's not surprising that Mary would learn something if *that* were all she knew".<sup>38</sup> Dennett imagines that Mary's captors play a trick on her when she leaves the black and white room and show her a blue banana. Since Mary has already determined exactly what kind of reaction she would have to each kind of color experience, by working out the details of her complete physical knowledge, she might just know that she is being tricked. Dennett's point is not that Mary would necessarily recognize the blue banana trick, but that defenders of Jackson's knowledge argument have no basis for claiming that she would not recognize the trick. The amount of knowledge that Jackson attributes to Mary inside the black and white room is so enormous that any claims about what Mary would or would not be able to determine on the basis of that knowledge is mere guesswork. As Dennett puts it, Jackson's knowledge argument is a "classic provoker of Philosophers' Syndrome: mistaking a lack of imagination for an insight into necessity".<sup>39</sup>

Similarly, Paul Churchland argues that if Mary indeed had a complete understanding of brain states she might be able to accurately imagine what any particular phenomenal experience would be like even if she had not yet had the experience. After all, she would apparently know exactly what kind of cortical activity would result from any possible stimulus. Churchland goes further than Dennett, however, in explaining *how* Mary might be able to figure out what it is like to see red merely on the basis of her completed neuroscientific knowledge. "Suppose that Mary has learned to conceptualize her inner life, even in introspection, in terms of the completed neuroscience we are to imagine".<sup>40</sup> Mary might describe to herself the inevitable loneliness she would feel inside

the black and white room in neuroscientific terms instead of in the emotive terms that you or I would use. In fact, Mary might know so much neuroscience that she would be able to determine what kind of mental experience would be associated with any given brain state. “If Mary has the relevant neuroscientific concepts for the sensational states at issue (viz., sensations-of-*red*), but has never yet been *in* those states, she may well be able to imagine being in the relevant cortical state, and imagine it with substantial success, even in advance of receiving external stimuli that would actually produce it”.<sup>41</sup> A real life parallel to how Mary might imagine what it is like to see red without having actually having seen red herself is the ability of highly trained musicians to imagine what a certain chord would sound like even if they had never heard that particular chord before. Similarly, when Mary first sees a red object she might, despite Jackson’s intuition otherwise, tell onlookers, “ho, hum – I already knew red objects would look like that”.<sup>42</sup>

Of course learning about color experiences by determining the experiential consequences of various specific neural states is, as Dennett points out, “not the usual way of coming to learn about colors, but Mary is not your usual person”.<sup>43</sup>

There are two reasons I consider the Dennett/Churchland account of what would happen to Mary upon leaving the black and white study an ineffective response to the knowledge argument. First, the suggestion that Mary would, inside the black and white room, be able to know what it is like to see red by imagining what it would be like to be in a particular brain state is counter-intuitive at best. The intuition that one must actually see red in order to know what a red visual experience is like is very strong. And the parallel to trained musicians imagining what a particular chord would sound like before

they hear it is weak. After all, musicians can imagine what a particular chord would sound like because they know what other chords sound like; however, Mary has been denied all chromatic experience. If the truth of physicalism depends on Mary's ability to recognize that the blue banana is a trick, then physicalism is in trouble.

Second, the Dennett/Churchland response misses the point. Even if Mary is able to imagine what a chromatic experience would be like on the basis of her complete physical knowledge the facts about what it is like to have chromatic experience would still be above and beyond the physical facts. As Jackson points out, "imagination is a faculty that those who *lack* knowledge have to fall back on".<sup>44</sup> If fact  $y$  can only be figured out (or imagined) on the basis of set of facts  $s$  then  $y$  is not part of  $s$ . If we accept Dennett and Churchland's analysis, the second premise of Jackson's knowledge argument ("After her release from the black and white room, Mary knows new facts") is indeed false, in that Mary learns nothing when she leaves her achromatic environment. However, the third premise ("There are facts other than all the physical facts") is still true. It's just that Mary also managed to learn (or imagine) those additional facts while still in the black and white room.

Nevertheless, Dennett and Churchland are right to point out that the amount of knowledge Mary is said to have inside the black and white room is so vast that we can only guess what Mary would and would not know. Current neuroscience offers little help in our guesswork. In the next hundred years our understanding of the brain will develop tremendously. If a top twenty-second century neuroscientist were to visit us, it is almost certain that we would understand little that she would say. And still, no one expects that neuroscience will be completed in the next hundred years. As Patricia Churchland points

out, Jackson is betting against the explanatory power of not just future science but of idealized, perfected, future science.<sup>45</sup> The history of science indicates that this is a bad bet. Perhaps Dennett is right and the second premise of the knowledge argument rests not on sound intuition but merely on our profoundly limited capacities of imagination.

*Jackson Post-1998: Mary acquires an illusion, not a fact, after leaving the black and white room.*

In 1998, Jackson renounced the knowledge argument and converted to physicalism.<sup>46</sup> It might seem as though Jackson's rejection of his own argument would put the discussion to rest once and for all. However, Jackson's own response to the knowledge argument is not terribly persuasive. Jackson says that the knowledge argument rests on the epistemic intuition that, "you cannot deduce, from purely physical information about us and our world, all there is to know about the nature of our world, because you cannot deduce how things look to us, especially in regard to color".<sup>47</sup> Though he continues to argue for the metaphysical claims of the knowledge argument, Jackson now contends that this epistemic intuition is based "on an illusion about the nature of color experience".<sup>48</sup> The illusion, as Jackson sees it, rests on a distinction between intentional properties and instantiated properties. Jackson formerly assumed that when we have a representation of a given object in qualitative experience, we have a representation of a real object or property. But this pre-philosophical assumption only establishes the existence of an *intentional* property; it does not demonstrate that the property is actually *instantiated* in the perceived object. For instance, when Mary leaves the black and white study and sees, say, a red rose she concludes that she has perceived

the red property instantiated in the rose; she assumes that the rose has the property of being red. Post-1998, Jackson rejects the assumption that phenomenal experiences necessarily represent some real property of the object represented. As he puts it, “the issue for us is whether the aspects that constitute the phenomenal nature of an experience outrun its representational nature, and there are good reasons to deny this”.<sup>49</sup> In other words, even though redness is an intentional property of the qualitative experience it is not an instantiated property of the object itself. In saying that phenomenal properties are intentional and not instantiated in the object, Jackson claims that phenomenal properties involve a “convenient, if metaphysically misleading, way of talking about how things are being represented”.<sup>50</sup> The critical claim is that the physicalist worldview need not accommodate the phenomenal character of our experiences if the intentional property of the phenomenal experience is not an instantiated property of the perceived object. If Mary believes that she learns something new about what it is like to see red then she is mistaken; she has learned about an illusion, not a fact. As Jackson puts it, “we may want to go so far as to say that sensing red misrepresents how things are. If this is right, we should say that nothing is red, for nothing would be as our experience of red represents things as being; we should be eliminativists about red and about color in general”.<sup>51</sup> At the very least, he suggests, we should hold that there are “complex physical properties” that we identify through color, even though the physical property is not itself color.<sup>52</sup>

I agree with Jackson that ‘redness’ is not itself an instantiated property of objects, and will grant for the sake of argument that, in some sense, we should be eliminativists about color. Redness is a property of the perception of the object, made possible by a combination of the physical properties of the object and the specific nature of our visual

processing system.<sup>53</sup> However I do not see why the controversy about whether redness is a real property of object bears on the knowledge argument. Earlier, Jackson had emphasized that what Mary did not know was a fact about other people, not a fact about the objects that people perceive.<sup>54</sup> If people experience a certain illusion when having chromatic experience then *it is still a fact about them that they are having the illusion*. Even granting eliminativism about color, it still seems to be the case that Mary acquires new factual knowledge when she leaves the black and white room; she learns the following fact about the world: “when people see red they are under *that* illusion”. Whether the property of redness is an instantiated property of the object is irrelevant; redness is still an instantiated property of the experience they are having – even if their experience is an illusion. Thus, Jackson’s claim that the physicalist worldview need not accommodate the phenomenal character of our experiences (because the intentional property of the phenomenal experience is not an instantiated property of the perceived object) ironically fails to take into account the clarification of what Mary did not know inside the black and white room that Jackson himself had made eighteen years earlier.

## II. Jackson Equivocates on ‘Knows’ (or ‘Physical’, or ‘Complete’)

Easily the most common physicalist response to the knowledge argument is that Jackson’s argument equivocates. That is, when Mary leaves the black and white room, she gains knowledge in a different sense of the word ‘knows’ (or ‘physical’, or ‘complete’) than the sense in which Jackson says that her pre-release knowledge was complete. Since the two references to ‘knows’ are not univocal in the first two premises, the sub-conclusion of line three of the argument simply doesn’t follow. Broadly speaking,



one of two strategies are employed by those who accuse Jackson of equivocating: asserting that 'knows' in the second premise doesn't amount to factual knowledge and asserting the facts known only upon release are also physical facts.

#### A. The Ability-Acquaintance Hypothesis

According to the ability-acquaintance hypothesis, when Mary leaves the black and white room, she acquires something other than factual knowledge.

*Lewis-Nemirow: Mary acquires know-how when she leaves the black and white room.*

The "ability hypothesis", most closely associated with David Lewis and Laurence Nemirow, is the most well known physicalist response to Jackson's knowledge argument.<sup>55</sup> Lewis suggests that, when Mary leaves the black and white room, she gains the ability to "remember, imagine, and recognize".<sup>56</sup> Only after she has her own chromatic experiences can she recognize red when she sees it, imagine a red experience without the presence of red objects, and remember what red things look like. Such abilities are a form of knowledge-how which is quite distinct from the knowledge-that which she knew inside the black and white room. Unlike knowledge-that, these abilities can not be taught through lessons. They can only be acquired through actually having the experience first hand. According to Lewis, physicalists must reject only that knowledge-how "represents a special kind of information about a special subject matter. Apart from that claim it is up for grabs what, if anything, it may represent".<sup>57</sup> Thus, Lewis is willing to grant that Mary knows every physical fact achromatically. He is also willing to grant that Mary would, in some sense, learn something new upon leaving her achromatic

environment. Additionally, he is willing to grant (if he must) that these abilities might be “a special kind of representation of some sort of information”.<sup>58</sup> However, if they are a representation of some sort of information, then that information is not information about a special (non-physical) property. Mary doesn’t learn *about* anything new when she leaves the black and white room; she just learns new abilities regarding the same entities and properties that she knew about all along. Since these abilities do not require the existence of non-physical entities or properties they are no threat to physicalism. In short, Lewis accuses Jackson of equivocation. The first and second premises are true only if they refer to different senses of *knowledge*. Before critiquing the ability hypothesis, it is useful to describe the acquaintance hypothesis.

*Conee: Mary acquires an acquaintance with certain facts when she leaves the black and white room.*

Earl Conee’s “acquaintance hypothesis” is an attempt to improve on the approach of the “ability hypothesis”. It rests on a distinction between factual knowledge and knowledge by acquaintance. Beyond the factual knowledge Mary perfects in the black and white room, and the abilities she acquires upon leaving the black and white room described by Lewis, Conee maintains that knowledge by acquaintance constitutes a third kind of knowledge. Knowledge by acquaintance “requires a person to be familiar with the known entity in the most direct way that it is possible for a person to be aware of that thing”.<sup>59</sup> For example, “to come to know a city is to become acquainted with the city, and to come to know a problem is to become acquainted with the problem”.<sup>60</sup> Though Mary acquires all of the factual knowledge regarding color experience inside the black and

white room she is not acquainted with color experiences herself until she leaves the black and white room. When Mary leaves she is able to think about phenomenal color experiences in a new way; in a sense it might even be said that she therefore gained new knowledge regarding color experience. But, as with Lewis's ability hypothesis, Mary does not come to learn *about* anything new. There is no threat to physicalism because there is no non-physical entity or property that Mary learns about only upon her release. This argument also claims that Jackson's argument equivocates. The first premise ("Inside the black and white room, Mary knows all the physical facts") is true only if 'knows' refers to factual knowledge and the second premise ("After her release from the black and white room, Mary knows new facts.") is true only if 'knows' refers to knowledge by acquaintance.<sup>61</sup>

*Paul Churchland: Mary acquires a pre-linguistic medium of representation when she leaves the black and white room.*

According to Paul Churchland's second objection to the knowledge argument, Mary's knowledge inside the black and white room amounts to mastering the kind of propositions that one finds in Neuroscience textbooks (though many, many more of them). Grappling for a precise description of the kind of knowledge Mary acquires only upon leaving her achromatic environment, Churchland writes that it "seems to be a matter of having a representation of redness in some prelinguistic or sublinguistic medium of representation for sensory variables, or to be a matter of being able to make certain sensory discriminations, or something along these lines".<sup>62</sup> In a later article, Churchland characterizes the difference between the two senses of *knowledge* in terms of a distinction

between knowing-how and knowing-that (like Lewis) as well as a distinction between propositional knowledge and knowledge by acquaintance (like Conee).<sup>63</sup> Thus, according to Churchland as well, Jackson's knowledge argument equivocates in that the second premise is true only in a different sense of 'knows' than the sense used in the first premise. Also like Lewis and Conee, Churchland maintains that the two senses of knowledge are two different ways to know *about the very same thing*.

Though Churchland is vaguer than either Lewis or Conee regarding the nature of Mary's post-release knowledge, his response is stronger than either of theirs in that Churchland demonstrates a double dissociation between the two relevant senses of knowledge by contrasting two people, one who has knowledge-that but not knowledge-how regarding a proper golf swing, and one who is in the opposite position.<sup>64</sup> Imagine a physics professor who teaches a course in the physics of sports but is not a good golfer herself. She would have a great deal of propositional knowledge, or knowledge-that, regarding a proper swing, but utterly lack any knowledge by acquaintance, or knowledge-how, regarding a proper swing. In contrast, imagine a skilled golfer who can not describe how to properly swing a golf club. Such a person would have great deal of knowledge-how, but very little knowledge-that regarding a proper golf swing. That both of these scenarios are entirely plausible (I know people who roughly meet each description) shows that the two kinds of knowledge are entirely independent. Even if Mary has a completed knowledge-that regarding chromatic experiences there is no reason to expect knowledge-how to follow.

### *Critique of the Ability-Acquaintance Hypothesis*

The problems with the above three accounts of Mary's post-release knowledge can be seen most clearly by examining Lewis's ability hypothesis. After pointing out the problems with Lewis's account, I hope to show that Conee's and Churchland's positions fare little better. First, knowing what it is like to have chromatic experiences does not necessarily provide one with the kinds of abilities that Lewis mentions. Suppose Mary undergoes the same surgery that HM did, bilaterally removing her hippocampus and amygdala, leaving her unable to form new memories.<sup>65</sup> She then leaves the black and white room for a short period of time before returning. That she would be unable to remember, imagine, or recognize what chromatic experiences are like does nothing to diminish the intuition that, while she was outside the black and white room, she knew what it is like to have such an experience. It is therefore implausible to reduce the epistemic event Mary has immediately upon leaving the black and white room to mental capacities that can be demonstrated only outside the context of the immediate experience itself. HM-Mary still has the striking chromatic experience that provides support for the intuitive second premise of the knowledge argument even though she gains none of the abilities Lewis mentions.<sup>66</sup>

Second, even if it is granted that Mary does acquire the abilities that Lewis mentions it is far from clear that she learns nothing but these abilities. It is not enough to argue that Mary gains something other than factual knowledge when she leaves the black and white room.<sup>67</sup> The claim must be made that she does not also gain factual knowledge, that she does not learn any additional truths about the world. And unless it can be shown that demonstrative facts are somehow not respectable, Jackson can still hold that Mary

learns (in addition to certain abilities) the following fact: “it is like *that* when people have red experiences”. If anything, the ability hypothesis seems to highlight the existence of even more facts that Mary can learn only upon release from the black and white room: “it is like *that* when people remember, recognize, or imagine red experiences”. Jackson’s own response to the ability hypothesis is along similar lines. He imagines that Mary is trying to decide whether an extreme form of solipsism is right. That is, whether there are even any other experiences beside her own. Jackson asks whether what she is to-ing and fro-ing about is an ability or, according to his view, a fact.<sup>68</sup> Considered in such a context, it seems implausible to suggest that the issue at hand is merely the possession of certain abilities. The ability hypothesis does little to undermine Jackson’s claim that Mary learns such truths about the world when she leaves the black and white room.

It seems more plausible that Mary does become acquainted with qualitative properties when she leaves the black and white room than that she gains certain abilities. Still, the claim that she acquires nothing but an acquaintance with certain facts that she already knew fares no better than the claim that she learns nothing but new abilities. That she gains knowledge by acquaintance also does nothing to undermine the claim that she also gains new factual knowledge. In fact, the ability-acquaintance hypothesis suggests yet one more true proposition that Mary can learn only upon leaving her achromatic environment: “being acquainted with/being able to recognize red is like *that*”. Given that Churchland’s description of Jackson’s equivocation adds little to the ability and acquaintance hypotheses, the problems with Lewis’s and Conee’s analysis also apply to Churchland’s.

It is not clear that Mary's post-release knowledge is *about* the same properties and entities she learned inside the black and white room. The very essence of Jackson's claim is that inside the black and white room Mary could not know *about* the subjective experience associated with chromatic experience. Lewis, Connee, and Churchland do little to establish that subjective experience does not constitute a subject matter in itself, distinct from the neurological processes that give rise to subjective experiences. As Lewis sees it, neurological processes do not "give rise" to subjective experiences because they *are* subjective experiences. But Jackson persuasively (to my mind at least) responds that when Mary is inside the black and white room she learns *about* brain states and processes involved in chromatic experience; when she leaves the black and white room she learns *about* subjective and qualitative chromatic experience. To me, it seems best to accept the straightforward language that there are facts *about* the subjectively accessed, qualitative aspect of experience that one could not know without having had qualitative experience of one's own. Put in a way that is neutral on the question of whether Mary learns about something new upon her release, the more persuasively it is shown that Mary's post-release knowledge should not be expected to follow from the kind of knowledge she has inside the black and white room (as Lewis, Connee, and Churchland argue) the less certain it is that the two kinds of knowledge are about the same thing (as Lewis, Connee, and Churchland also claim).

The strength of the ability-acquaintance hypothesis is that it acknowledges the rich nature of qualitative experience and the importance of providing a physicalist account of qualitative experience. Interestingly, most of Lewis's article on the knowledge argument explains six different ways to miss the point of Jackson's argument. Though I

think Lewis overstates the extent to which Jackson's knowledge argument is a threat to physicalism, by acknowledging that experience is the only practical way to know what a particular qualitative state is like he contributes to overcoming the traditional physicalist tendency to ignore or eliminate the epistemic value of subjectivity. In short, the strength of the ability-acquaintance hypothesis is that it acknowledges Jackson's epistemic intuition; Lewis acknowledges that chromatic experiences provide information that can be acquired no other way.

I consider it a weakness of the ability-acquaintance hypothesis that the image of Mary knowing (in some relevant sense) all of the physical facts in black and white is unchallenged, as is the notion that physicalism is committed to the completeness of Mary's black and white factual knowledge. The intuition that Mary's complete physical knowledge leaves something out is also unchallenged. Jackson's ambiguous use of 'physical', and physicalism's commitment to completeness are unexplored by advocates of the ability-acquaintance hypothesis. Since Lewis, Conee, and Churchland are all reductive materialists, it is not surprising that they fail to challenge Jackson's assumption that physicalism must be reductive, but the end result is a less than compelling response to the knowledge argument.

#### B. The Differently Physical Facts Hypothesis

Whereas proponents of the ability-acquaintance hypothesis argue that Mary gains something other than factual knowledge upon release from the black and white room, proponents of the differently physical facts response (Musacchio, Horgan, and Alter) argue that Mary's post-release knowledge is a different kind of physical-factual



knowledge. The *differently physical facts* response argues that Mary's knowledge inside the black and white room is complete only in regards to one type of physical knowledge. Though Mary learns new facts when she leaves the black and white room, physicalism is consistent with the existence of such facts. Proponents of the ability-acquaintance hypothesis argue that Jackson equivocates on two different senses of knowledge in the first two premises; proponents of the differently physical facts hypothesis argue that Jackson equivocates on two different senses of the complete physical facts in the first and fourth premises.

*Musacchio: When Mary leaves the black and white room she gains (physical) phenomenal knowledge.*

Jose Musacchio offers the most direct version of the differently physical facts response. Musacchio offers a neuro-biologically grounded distinction between “propositional knowledge” and “phenomenal knowledge”. By ‘propositional knowledge’ Musacchio means knowledge that can be expressed in propositions. Propositional knowledge is a higher brain function, dependent on the capacity of the cortex to support language.<sup>69</sup> By ‘phenomenal knowledge’ Musacchio means “the collection of phenomenal experiences and phenomenal concepts” that an individual possesses about the world.<sup>70</sup> Phenomenal knowledge is a function of phylogenically old brain structures. Unlike a particular proposition that I can share with you (assuming we share a language, are both educated enough to understand the proposition, etc.), I can not share my phenomenal concepts. Propositional knowledge is public; phenomenal knowledge is private. No matter how hard I might try to explain what ‘red’ is to blind persons I can

never make them understand. In fact, according to Musacchio, at least some non-linguistic animals possess and understand phenomenal knowledge, whereas propositional knowledge is possessed exclusively (as far as we can tell) by human beings.<sup>71</sup> Thus, Musacchio's position is similar to Paul Churchland's critique that Mary knew discursive knowledge in the black and white room. The difference is that, unlike Musacchio, Churchland argues that Mary's post-release knowledge is knowledge about the same things as her pre-release knowledge.

While Mary is locked in her black and white study, she comes to know all of the propositional knowledge regarding visual processing that there is to know. But her vast propositional knowledge provides her with no phenomenal knowledge regarding color. She can acquire the latter only upon leaving the black and white room. Therefore Jackson's thought experiment establishes only that complete propositional knowledge is not complete knowledge *simpliciter*; there is a kind of knowledge that can not be expressed propositionally.<sup>72</sup> Musacchio offers a summary of the emerging neurobiological explanation of phenomenal knowledge to establish that such phenomenal knowledge is nonetheless physical.

In short, Musacchio's position is that the knowledge argument equivocates in that the first premise ("Inside the black and white room Mary knows all the physical facts") is true only if 'physical facts' refers to propositional knowledge; whereas, the second premise ("After her release from the black and white room Mary knows new facts") is true only if 'new facts' refers to phenomenal knowledge. By providing a neurobiological account of phenomenal knowledge, Musacchio focuses attention on the fourth premise of the knowledge argument and the sense in which physicalism is committed to

completeness by arguing that genuine phenomenal factual knowledge is physical.

Though Musacchio makes a valuable contribution to the discussion of the knowledge argument by providing (albeit somewhat speculatively) a neuro-biological account of Mary's post-release knowledge, I think that Musacchio's response to the knowledge argument does not work. That Mary's post-release epistemic events are possible only because of particular brain structures of the human brain is common ground for defenders and critics of the knowledge argument. Jackson is categorically *not* claiming that Mary's post-release factual knowledge is possible because of something like an immaterial soul. The question is whether the truths Mary acquires when her phylogenically old brain structures become active for the first time are physical truths. Defenders of the knowledge argument will grant Musacchio's neurobiological analysis of the basis of qualitative experience and nevertheless assert that phenomenal knowledge is knowledge *about* the non-physical properties of human experience. That particular brain structures are what make qualitative facts knowable does not imply that such facts are themselves physical facts about the brain.

*Horgan: When Mary leaves the black and white room, she learns facts that are ontologically but not explicitly physical.*

According to Terence Horgan, inside the black and white room Mary acquires all of the *explicitly physical facts*, that is, she acquires every fact that "belongs to, or follows from, a theoretically adequate physical account".<sup>73</sup> She does not necessarily possess all of the *ontologically physical facts*.<sup>74</sup> A proposition expresses an ontologically physical fact "just in case (i) all the entities referred to or quantified over in [the proposition] are

physical entities, and (ii) all the properties and relations expressed by the predicates of [the proposition] are physical properties and relations”.<sup>75</sup> Whereas explicitly physical facts are, by definition, expressed in “overtly physicalist language”, ontologically physical facts “can be expressed by other sorts of language – for instance, mentalistic language”.<sup>76</sup> In other words, the domain of ‘facts about physical things’ is larger than the domain of ‘facts that can be expressed in explicitly physical language’.<sup>77</sup> Some facts, precisely the kind Mary learns only upon release from the black and white room, are ontologically physical but are not expressed in overtly physicalist language. Since physicalism is only committed to the assertion that all facts are ontologically physical, the fact that Jackson’s thought experiment demonstrates there are facts other than explicitly physical facts is no threat to physicalism. Horgan characterizes his critique by arguing that the knowledge argument equivocates: the first premise (“inside the black and white room Mary knows all the physical facts”) is true only if ‘physical facts’ refers to explicitly physical facts; whereas, the fourth premise (“if physicalism is true then there are no facts other than the physical facts”) is true only if ‘physical facts’ refers to ontologically physical facts.<sup>78</sup>

I ultimately agree with Horgan that there are facts about physical things that are not expressible in the terminology of physics, chemistry, or biology, though he could do more to explicitly make the connection between the sense in which physicalism claims completeness and the sense in which Mary’s knowledge is asserted by Jackson to be complete. Nevertheless, Horgan’s response to the knowledge argument has a strategic weakness. Even if it is granted that ontologically physical information *need not be* expressed in physical language, the question remains whether Mary’s post-release

knowledge *in fact* refers only to physical properties and entities. Defenders of the knowledge argument might respond that at least some of the properties relevant to Mary's post-release knowledge are not physical; Horgan offers virtually nothing to rebut such an assertion. One might grant his argument that there are ontologically physical facts that are not explicitly physical (such as, perhaps, economic facts) and still maintain that when Mary is released she learns facts that are neither explicitly nor ontologically physical. Jackson could respond to Horgan by saying that, *ex hypothesi*, Mary knew all of the ontologically physical facts inside the black and white room. And by granting that Mary's post-release knowledge consists of facts that are not explicitly physical, Horgan will be hard pressed to justify the claim that such facts are, in fact, ontologically physical. If they don't look like physical facts, there is little to ground the claim that they actually are physical facts.<sup>79</sup> One could go further and suggest that Horgan's distinction between ontologically physical facts and explicitly physical facts creates *prima facie* doubt about whether qualitative facts ought to be considered physical. In contrast to the view I defend in Chapter Three that qualitative facts can have scientific respectability (too complicated to quickly summarize here, you will just have to wait), I think Horgan goes too far by conceding that qualitative facts are not explicitly physical.

*Alter: When Mary leaves the black and white room she learns facts about physical things that can not be learned through discursive means.*

Torin Alter's response to the knowledge argument is quite similar to Horgan's, but is cast in better terms. Whereas Horgan refers to "explicitly physical facts", Alter refers to "facts learned through discursive means"; whereas Horgan refers to

“ontologically physical facts”, Alter refers to “facts about physical things”.<sup>80</sup> That is, Alter grants that while Mary is still in the black and white room she knows all the facts that can be learned discursively, but that upon release she learns additional facts about physical things.<sup>81</sup> Unlike Horgan, however, Alter explicitly casts the discussion in terms of the sense in which physicalism claims completeness. Specifically, Alter agrees that the knowledge argument refutes any version of physicalism that accepts the following proposition (which he names DL): “all facts about physical events are discursively learnable”.<sup>82</sup> He further argues that most existing versions of physicalism in fact accept DL – and are therefore refuted by Jackson’s knowledge argument. However, Alter argues that physicalism *need not* accept DL; it is by mere historical accident that most existing versions of physicalism do accept DL. According to Alter, the lesson to be drawn from the knowledge argument is that versions of physicalism need to be developed that are not committed to DL.

The greatest strength of Alter’s response is that it shifts the focus from whether physicalism is true to the conditions for the possibility of physicalism in light of the intuitive strength of Jackson’s thought experiment. The net effect is that his use of ‘facts learned through discursive means’ instead of Horgan’s ‘explicitly physical facts’ largely mitigates the critique I previously leveled against Horgan. There is not the same *prima facie* doubt regarding the physical credentials of ‘facts about physical things which are not discursively learnable’ as there is regarding ‘ontologically physical information which is not explicitly physical’. It is more plausible to suggest that physicalism is committed to the completeness of the explicitly physical facts than to suggest that physicalism is committed to the completeness of discursively learnable facts. After all, it

is quite reasonable to suggest that there are non-scientific, yet discursively knowable, facts.

However, in mitigating one part of the critique I leveled against Horgan, Alter makes himself more susceptible than Horgan to a part of that critique which might be leveled against any version of the differently physical facts response. Jackson might reply that the first premise of the knowledge argument establishes, *ex hypothesi*, that Mary has *all physical knowledge* not just all *explicit physical knowledge* or *all discursively learnable knowledge* or *all propositional knowledge*. Jackson can then respond that, if Mary knew only the discursively learnable physical facts while she was inside black and white room, we should not be surprised that she learns additional physical facts.

However, Jackson might continue, that it is simply not the argument that he was making: Mary knows all the physical facts that there are – whether they be discursively learnable, whether they be explicitly cast in physical language. Because such a response by Jackson would clearly play on his persistent ambiguity about the meaning of ‘physical’ and the resulting ambiguity about what Mary is exactly alleged to know inside the black and white room, I don’t think Jackson’s response ultimately works. Nevertheless Mussachio, Horgan, and Alter take advantage of the same ambiguity when they provide their own accounts of what Mary knew inside the black and white room. The strategic advantage afforded to Jackson by his ambiguous use of ‘physical’ can be nullified only if physicalist responses to the knowledge argument insist, as the logic of the argument demands, that Mary is only possible if it is possible to know all the facts *as far as physicalism is concerned* in a black and white room. That is, while I find little in Alter’s or Horgan’s positions on the knowledge argument to disagree with, I think that focusing on the first

premise of the knowledge argument (and the implications of Jackson's *ex hypothesi* stipulation) is a far better way for physicalists to cast the debate than to fill in for Jackson by providing a specific account of what Mary knows inside the black and white room.

### III. Same Facts, New Mode of Access

*Loar: When Mary gets out of the black and white room, she learns new concepts of the same properties that she learned about before her release.*

According to Brian Loar's response to Jackson's knowledge argument, "the physicalist should accept Jackson's intuitive description of Mary: she fails to know that we have certain color experiences even though she knows all relevant physical facts about us. And when she acquires color experience, she does learn something new about us – if you like, learns a new truth or fact".<sup>83</sup> Having granted that Mary learns something new and irreducible to physical facts upon her release, Loar argues that the knowledge argument rests on an unstated semantic premise: "a statement of property identity that links conceptually independent concepts is true only if at least one concept picks out the property it refers to by connoting a contingent property of that property".<sup>84</sup> If the semantic premise is false then it is possible that Mary's pre- and post-release knowledge are conceptually independent, and equally essential, understandings of the same properties. That is, Loar's analysis centrally features a distinction between properties and concepts. Concepts pick out properties and, on Loar's view, phenomenal concepts and physical functional concepts pick out the same properties. Loar accepts that phenomenal concepts are irreducible in that, "they neither a priori imply, nor are implied by, physical-functional concepts".<sup>85</sup> It would not be possible for Mary to learn the phenomenal



concepts based merely on exhaustive knowledge of the physical-functional concepts. Statements such as 'Mary knows physical property x under such and such physical-functional concepts' do not maintain their truth value when 'phenomenal concepts' replaces 'physical-functional concepts'.

Anti-physicalists like Jackson (pre-1998) assume this conceptual irreducibility implies that phenomenal *properties* are not reducible to physical *properties*. That is, if the concept is irreducible then the property that it is a concept of must also be irreducible.<sup>86</sup> According to Loar, there is no grounding for this assumption. Phenomenal concepts are a kind of recognitional concepts that refer to complex physical and functional properties of the brain. While Mary is inside the black and white room she learns every physical concept that there is to be known about the relevant physical-functional properties of the brain, and when she is released she learns about those same physical-functional properties of the brain *from an experiential point of view*. The same thing can be known under two different modes of knowing. And, as the phenomenal and physical-functional concepts both refer to the same physical properties, statements about qualitative experience cast in either conceptual language can refer to identical properties of the brain.

Thus, on Loar's view, the knowledge argument is best understood as a claim that there are irreducible phenomenal concepts and a claim that, if physicalism is true, then physical properties can be characterized only by physical concepts. Loar points out the similarity between the semantic premise and Kripke's account of identity statements, and suggests that Jackson's and Kripke's ant-physicalist arguments are related. For both Kripke and Jackson, physicalism fails because physical concepts (and not phenomenal concepts) seem to rigidly designate physical properties of the brain; and, to the extent that

phenomenal concepts designate the physical properties of the brain, they must therefore do so only contingently. Loar's position is not a version of the differently physical facts hypothesis in that he is not suggesting that there are physical properties above and beyond those that can be learned in a black and white room, but that there are modes of epistemic access to physical properties that are not available achromatically. Unlike the differently physical facts response, Loar holds (counter-intuitively, it seems to me) that Mary only seems to be learning new facts about the world because she is actually learning the same old facts under a new epistemic presentation. She acquires additional knowledge, but only in the sense that she knows some things now also "in the mode of experience". As Loar put it, "what she lacked and then acquired, rather, was knowledge of certain such properties couched in experiential terms".<sup>87</sup>

I think that Loar's analysis raises serious questions about the physicalist credentials of Mary's post-release knowledge that are similar to the strategic weakness that I suggested arise when physicalists accuse Jackson of equivocation (especially, see my critique of Horgan). Given that phenomenal concepts are neither reducible to physical concepts nor expressible in explicitly physical language, there is little evidence for Loar's claim that phenomenal concepts refer ultimately to physical-functional properties. In fact, by acknowledging that phenomenal concepts can not be expressed in physical terms, Loar has denied the possibility of any explicit evidence for the claim that phenomenal concepts refer to physical properties. Consider the resulting relative standpoints of Loar's position and defenders of the knowledge argument. They agree that there are phenomenal concepts that are not reducible to the physical concepts. They agree that the physical concepts capture physical properties essentially. In the absence of compelling evidence to

the contrary, the most natural supposition is that the phenomenal concepts capture phenomenal properties essentially. It is counter-intuitive, after all, to hold that there is a single property that can be essentially referred to by both 'what it is like to see red' and 'x neural pattern'.

Furthermore, Loar's willingness to bite this bullet is, arguably, *ad hoc*. Chalmers more or less claims that Loar will hold onto an ontologically simplifying identity thesis at all costs.<sup>88</sup> Though Loar's epistemological views are non-reductive in the sense that phenomenal concepts do not retain their truth value when replaced by physical concepts, again and again his article highlights that it is nevertheless possible to hold that phenomenal states are identical to physical-functional states of the brain. In light of Loar's acceptance of Jackson's claim that there are facts that can not be reduced to physical concepts, other than a willingness to turn to anything that might rescue physicalism, it is hard to see what motivates the view that such phenomenal facts nevertheless refer to physical properties. Loar's argument would be much stronger if he were to show that there are any facts (even facts other than phenomenal facts) that are identical to, but not reducible to, physical facts.<sup>89</sup>

Chalmers claims that Loar hasn't made the case that there is only one property, that there are only dogmatic reasons for Loar to hold on to an ontologically simplifying version of physicalism, and that the core epistemological intuition that he and Loar share is more effectively expressed as a form of property dualism. Judging from where I'm sitting, Chalmers wins the debate.

Ironically, Chalmers might argue that the response to the knowledge argument that I defend in Chapter Two (and the speculative view proposed in the Conclusion) goes

further down the road to property dualism even than Loar's position. It seems to me that it is best to acknowledge that there are qualitative properties of experience that are distinct from the properties known by means of physics, chemistry, and biology. Though I reject ontologically simplifying accounts of phenomenal facts and properties, I think that such phenomenal properties are nevertheless physical properties in the sense that is relevant to the completeness of physicalism. As I see it, the value of the knowledge argument lies precisely in recognizing the need to understand the sense in which physicalism can claim to be complete while recognizing the qualitative *properties* of lived experience.

## Conclusion

Virtually all of the many physicalist responses to the knowledge argument focus on what happens when Mary *leaves* the black and white room. Almost no one challenges the image of Mary *inside* the black and white room. Though I have argued that none of the conventional physicalist responses effectively answers Jackson's knowledge argument, I neither accept Jackson's anti-physicalist conclusion nor believe that physicalists have no effective response available. The problem with the conventional physicalist response is that they acknowledge that Mary's knowledge inside the black and white room is complete *in any sense whatsoever*. Once it is conceded that there is a complete set of facts *s* that are knowable achromatically, *s* is established as the privileged kind of physical knowledge. Assertions that any other kind of facts deserve to be called physical are automatically suspect. Allowing that Mary's post-release knowledge is outside the privileged set of facts *s* grants Jackson a strategic victory that is, due to

Jackson's ambiguity about what constitutes a *physical* fact, difficult to overcome. In response, Jackson can take advantage of his ambiguity about the meaning of 'physical' and claim that, in addition to any other unique aspect of her post-release life, Mary learns more facts that properly belongs to *s* (as in his response to Lewis and Conee). Alternatively, Jackson can again take advantage of his ambiguity about the meaning of 'physical' and claim that Mary's pre-release knowledge is broader than *s* (as in his response to Horgan and Alter). In either event, physicalist attempts to clarify what Mary knew inside the black and white room, and what she learned upon release, ultimately provide strategic advantage to defenders of the knowledge argument.

Instead of focusing on what happens when Mary leaves the black and white room, physicalists ought to highlight Jackson's persistent ambiguity regarding the meaning of 'physical facts' and explore directly the sense in which physicalism is actually committed to completeness. In the end, physicalists should patently refuse to discuss what happens when Mary leaves the black and white room and instead point out that Jackson isn't even clear about what happens while Mary *is still inside*. It does not help physicalists to pretend that they understand precisely what Mary knows before she leaves her study, however plausible the account. Physicalists are better off if they refuse to play the game, and outright dismiss the possibility of a physically omniscient achromatic knower.

Though Dennett chides other philosophers for not resisting Jackson's image of Mary, the image that Dennett refers to is that of Mary learning something new *when she leaves* the black and white room.<sup>90</sup> No one challenges the image of Mary *inside* black and white room. Denying that it is possible (even in principle) to know all the physical facts

inside a black and white room 1) gets physicalists what they want, and 2) avoids some of the problems with the other strategies.

---

#### Notes to Chapter One

37 Dennett. "Epiphenomenal Qualia?," 60. See also Dennett's more recent discussion of Mary and the blue banana in Dennett, *Sweet Dreams*, 103-129. Though Dennett's discussion is updated to reflect recent literature, he continues to focus his critique on Jackson's claim Mary will "surely" learn something when released.

38 Ibid, 60.

39 Ibid, 62.

40 Churchland. "Reduction, Qualia, and the Direct Introspection of Brain States," 25.

41 Ibid, 26.

42 Jackson says that Mary will not say "ho hum." See Jackson, "What Mary Didn't Know," 55.

43 Dennett, "Epiphenomenal Qualia?," 62.

44 Jackson, "What Mary Didn't Know," 52.

45 Patricia Churchland, *Neurophilosophy* 333.

46 I find different dates referred to in the literature for Jackson's renunciation of the knowledge argument, but 1998 is most commonly cited.

47 Jackson, "Mind and Illusion," 422.

48 Ibid, 426.

49 Ibid, 429.

50 Ibid, 427.

51 Ibid, 432.

52 Jackson further suggests that if we take this latter approach we should accept Lewis's ability hypothesis as a way to characterize our chromatic identification of complex physical properties. In other words, Jackson concludes that either the ability hypothesis or eliminativism is the best physicalist response to the knowledge argument and which of those two we choose will be determined by whether there is any instantiated property behind the intentional property of chromatic experience.

53 In the background here is the controversy regarding what Block calls *mental paint* (the existence of inner representations); see Block, "Mental Paint". Whereas Jackson consistently acknowledges the existence of such inner representations, post-1998 he believes that such representations outrun the instantiated properties of the object.

54 Jackson, "What Mary Didn't Know," 52.

55 The ability hypothesis is also sometimes referred to as the Lewis-Nemirow hypothesis in reference to Lewis' debt to Nemirow's "Review of *Mortal Questions* by Thomas Nagel," 473-477.

56 Lewis, "What Experience Teaches," 102.

57 Ibid, 102.

58 Ibid, 102.

59 Conee, "Phenomenal Knowledge," 207.

60 Ibid, 202.

61 The acquaintance hypothesis is also defended by Michael Tye. Though Conee's account of the acquaintance hypothesis is generally more clear than Tye's, Tye goes further than Conee in critiquing the ability hypothesis in that he demonstrates a double dissociation between Lewis's proposed abilities and knowing what it is like to have a particular qualitative experience. See Tye, "Knowing What it is Like," 154. For yet another version of the acquaintance hypothesis see Bigelow and Pargetter, "Acquaintance with Qualia," 179-195.

62 Paul Churchland, "Reduction, Qualia, and Direct Introspection of Brain States," 23.

63 Churchland, "Knowing Qualia: A Reply to Jackson," 164-165. Though he cites both Conee's and Lewis' formulations he also indicates a preference for Conee's more precise formulation.

64 Both Conee and Lewis also assert that we should not expect the kind of knowledge Mary has in the black and white room to also provide her with the kind of knowledge she acquires only upon leaving, but

---

neither Lewis nor Conee go so far as to demonstrate a double dissociation between the two kinds of knowledge.

65 Many thanks to graduate student colleagues in a seminar on Consciousness at Michigan State University for pointing out this critique.

66 The argument made in this paragraph is also made quite persuasively in Alter, "A Limited Defense of the Knowledge Argument," 37.

67 Jackson makes this point in "What Mary Didn't Know," 52.

68 Jackson, "What Mary Didn't Know," 55.

69 Musacchio, "Dissolving the Explanatory Gap," 333.

70 Ibid, 336.

71 I do not agree with Musacchio's assessments of the propositional capacities of non-human animals. I tend to doubt that there is any particular cognitive functioning that will turn out to present exclusively in human beings. Nevertheless, Musacchio's characterization of propositional language as a characteristically human attribute both helpfully illuminates Musacchio's position and, interestingly, resonates with the philosophical tradition dating back at least to Aristotle.

72 For a slightly different response to the knowledge argument based on the ineffable nature of Mary's post-release knowledge see Hellie, "Inexpressible Truths and the Allure of the Knowledge Argument," 333-364.

73 Horgan, "Jackson on Physical Information and Qualia," 304.

74 Actually, Horgan uses the word "information" instead of fact – but only while noting that it is inadvisable to think of bits of information as entities; his stated intention is to set such issues to the side in order to focus on why he thinks the knowledge argument fails. Since he was writing before Jackson's 1986 clarification that what Mary didn't know inside the black and white room, in adopting the terminology of "information" Horgan is merely adopting Jackson's terminology *at the time*. Since I have consistently used 'fact' throughout the dissertation, it seems only fair to also update Horgan's adoption of Jackson's terminology.

75 Ibid, 304.

76 Ibid, 304. Interestingly, Horgan's claim that ontologically physical information can be expressed in language other than explicitly physical terms opens a much wider range of possibilities than even Horgan (apparently) considers. Most dramatically, it leaves open the possibility of theological language being used to express physical information. The term 'god' might be the only effective way to capture extraordinarily complex physical properties of the world.

77 Virtually the same distinction is made by Daniel Stoljar, who claims that Jackson equivocates between two different senses of 'physical'. The first sense of physical is the "theory-based conception" in which a physical property is a property "that physical theory tells us about". The second sense of 'physical' is the "object-based conception" in which a physical property is either a property of the theory based conception or a property that supervenes on the properties of the theory based conception. See Stoljar, "Two Conceptions of the Physical," 312.

78 Contrast Horgan's account with other descriptions of Jackson's equivocation: because Horgan distinguishes between Mary's pre-release knowledge and a broader category of knowledge which also includes the knowledge she can acquire only upon release from black and white study, the two relevant premises for his characterization of Jackson's equivocation are the first and fourth premises instead of the first and second premises. That is, Horgan's (and Alter's, and Loar's) response to the knowledge argument express a disagreement with Jackson about the sense in which physicalism is committed to the completeness of the physical facts. In this way, they are more similar to the response I advocate in Chapter Two than other responses. Nevertheless, both Horgan and (to a lesser extent) Alter address the question of what happens to Mary when she leaves the black and white room.

79 I am grateful to my Grand Valley colleague Ronald Loeffler for pointing out this strategic weakness in Horgan's account.

80 Alter, "A Limited Defense of Knowledge Argument," 50.

81 One notable difference between the two is that, unlike Horgan, Alter does not directly claim that the knowledge argument involves an equivocation. Nevertheless, it is not hard to see how Alter's position can be expressed as an equivocation charge. The first premise of the knowledge argument is true only if "all the

---

physical facts” means “all discursive knowledge”. The fourth premise is true only if “all the physical facts” means “all the facts about physical things”.

82 Ibid, 51.

83 Loar, “Phenomenal States (Revised),” 222.

84 Ibid, 224

85 Ibid, 220.

86 Their positive arguments for the position are discussed in Chapter Four.

87 Ibid, 222.

88 Horgan also claims that qualitative states are identical to brain states. I present the current critique as a critique of Loar, and not of Horgan, because one need not hold to an ontologically simplifying relationship between qualitative states and physical-functional brain states in order to accept Horgan’s response to the knowledge argument.

89 For such an approach used in response to Jackson, see the discussion of Block and Stalnaker, “Conceptual Analysis, Dualism, and the Explanatory Gap” in Chapter Four.

90 Dennett, “Epiphenomenal Qualia?,” 60.



## Chapter Two:

### Zombie University

Denying the logical possibility of a physically omniscient but achromatic knower such as Mary is the best physicalist response to Jackson's knowledge argument. I do not mean there are mundane practical problems associated with locking someone up for most of their life and then having them suddenly see a vibrant shade of red.<sup>91</sup> I mean that physicalists should reject the following modal intuition embedded in Jackson's argument: it is possible for Mary to know all the physical facts inside a black and white room in the same sense of 'all the physical facts' in which physicalism claims that there are no facts other than all the physical facts. If the term 'all the physical facts' designates all the facts that can be placed in a physicalist ontology, then there are physical facts that can not be learned in a black and white room. Instead of answering the question of what happens when Mary is released from the black and white room, physicalists ought to argue that Jackson asserts an impossibility when describing what happens before Mary's release. I liken Jackson's modal intuition that it is possible to know all the physical facts achromatically to Chalmers' modal intuition that zombies are logically possible and argue that it is metaphysically prudent to reject such intuitions.

#### I. The Two Parts of the Knowledge Argument

As the knowledge argument is usually understood there are two claims that Jackson needs to prove. The first claim is epistemological and is supported by the thought experiment about Mary. As Jackson summarizes his epistemological claim, "you cannot

deduce, from purely physical information about us and our world, all there is to know about the nature of our world, because you cannot deduce how things look to us, especially in regard to color”.<sup>92</sup> The second claim is metaphysical, asserting that if physicalism is true then all information can be deduced from physical information. Though this metaphysical part is glossed over in Jackson’s articles that concern Mary, it is more developed in his recent work on conceptual analysis. As explained in the Introduction, I take it that the following formulation of Jackson’s argument clearly expresses Jackson’s intention:

### The Knowledge Argument

1. Inside the black and white room Mary knows all the physical facts.
2. After her release from the black and white room Mary knows new facts.
3. Thus, there are facts other than all the physical facts.
4. If physicalism is true, then there are no facts other than all the physical facts.
5. Therefore, physicalism is false.

As the knowledge argument is usually understood, the epistemological sub-argument is found in lines one through three. The metaphysical sub-argument is found in lines three through five. Jackson needs both the epistemological claim and the metaphysical claim in order to establish his anti-physicalist conclusion. A critical consequence of seeing the argument this way is that the thought experiment about Mary is seen as establishing a purely epistemological conclusion. The question about what Mary knows is considered distinct from the question about the commitments of physicalism. And Jackson

inadvertently reinforces this way of seeing the argument by discussing the sense in which physicalism is committed to completeness in articles where he largely ignores Mary, and vice versa.

I will argue that this way of seeing the relationship between metaphysics and epistemology in the knowledge argument glosses over important metaphysical aspects of Jackson's formulation of his epistemological intuition and thought experiment about Mary. That is, seeing the argument in the usual way under-appreciates the crucial connection between the first and fourth premise. The first premise stipulates that Mary knows all the physical facts inside the black and white study. The fourth premise says that if physicalism is true then there are no facts other than all the physical facts. While there is certainly some sense in which premise four is true, the knowledge argument only gains traction because Mary is said to have all of the physical facts in the sense of 'physical fact' in which physicalism claims that there are only physical facts. It is therefore an oversimplification to suggest that the thought experiment about Mary supports a purely epistemological conclusion. The thought experiment itself is infused with the metaphysical notion of the complete set of physical facts and, as a result (or perhaps it is the cause), Jackson's description of his epistemological intuition is infused with the metaphysical notion of 'purely physical information'. Though premise four is undeniably true *in some sense* (physicalism claims to offer a complete picture of the world) it generally goes unnoticed that, as Jackson sets up the argument, Mary is said to have a complete knowledge *as far as physicalism is concerned*. Discussions of the relationship between epistemological and metaphysical issues in the knowledge argument

rarely reflect the metaphysical implications of stipulating that Mary knows all the physical facts inside a black and white room.

I do not deny that the knowledge argument can be separated into distinct epistemological and metaphysical parts. I will only argue that the epistemological part of the knowledge argument is narrower than is generally acknowledged, and that the metaphysical part is broader than is generally acknowledged. Only when the narrow epistemological intuition at the heart of Jackson's argument is isolated and stripped down to its epistemological core can Jackson's metaphysical intuitions be properly understood. Once Jackson's epistemological and metaphysical intuitions are separated and stated more precisely, both the intuitive strength of the knowledge argument and the failure of the knowledge argument to defeat physicalism will be clarified.

## II. Jackson's Two Intuitions

Whereas Jackson acknowledges that the knowledge argument rests on a single intuition, I find two. The first is that there are facts that can only be known chromatically. The second is that such chromatically known facts can not be placed in a physicalist ontology.

### *The epistemological intuition*

The knowledge argument's epistemological intuition is expressed in the second premise: Mary learns new facts when she first has her own chromatic experiences. Jackson wants to cast this intuition in terms of the existence of facts that can not be deduced from the physical facts. However, to do so conflates metaphysical and

epistemological issues. Whether there are facts that can only be known chromatically is one question; whether such facts deserve to be called physical is a separate question. Jackson's thought experiment became an instant classic because of its compelling image of Mary's leaving her black and white room and seeing a red rose for the first time. Pre-reflective intuition strongly suggests that she will have an exciting and informative experience. But the reason it is just obvious that "Mary learns new facts when she leaves the black and white room" (premise two of Jackson's knowledge argument) is that she was denied chromatic experience inside the black and white room; that she also knew all of the physical facts is a stipulation of the first premise that is irrelevant for the intuition that her first chromatic experience will be interesting and informative. The epistemological intuition is just as effectively stated by imagining Harry, an otherwise ordinary person who spends his entire life inside a black and white room. Imagine that Harry isn't particularly curious; there is nothing especially interesting about what he does or does not know. Claiming that there are facts about chromatic experience that Harry *could never know* captures the intuition no less than imagining the physically omniscient Mary leaving the black and white study. Put bluntly, in order to capture Jackson's epistemological intuition no reference to 'physical' is necessary.

It is also worth noting that Jackson claims both that 1) there are facts that are *known only through chromatic experience*, and 2) such facts *describe properties of chromatic experience*. One might argue that Mary's story can only establish the existence of facts that can only be known through chromatic experience, but that since such facts are ultimately about complex physical properties, they represent little threat to physicalism (a strategy similar to the "differently physical facts" response discussed in

Chapter One).<sup>93</sup> I do not intend to make such an argument; though the distinction between qualitative experience as a method for learning and qualitative experience as a subject matter in itself is relevant for understanding certain points I will discuss later. I highlight it now merely in order to clarify Jackson's epistemological intuition and to state it as precisely as I can, stripped of all references to counterfactuals. I take it that the following is an accurate statement of the epistemological intuition that is at work in the knowledge argument:

Jackson's Epistemological Intuition: chromatic experience is the only possible method for learning certain facts about the qualitative properties of chromatic experience.

There has been exhaustive debate about whether Mary's learning should be characterized in terms of learning new "facts".<sup>94</sup> It is an issue I do not take up in this chapter, as it is tangential to Jackson's core epistemological intuition that there is *some kind of rich epistemic event* that can only take place during lived chromatic experience. Jackson wants to call this rich epistemic event "learning a new fact" and that seems as good as any way of putting the intuition. No part of my response to the knowledge argument hinges on whether we call the uniquely informative epistemic event associated with chromatic experience "learning a fact about the world". For the sake of consistency and clarity, I reserve the phrase 'qualitative fact' to refer to those facts that Jackson's epistemological intuition asserts the existence of.

I enthusiastically endorse Jackson's core epistemological intuition knowing full well that most physicalists will think I have cluttered our desert landscape of scientifically enlightened truths with introspection and subjectivity. Nevertheless, endorsing Jackson's intuition takes consciousness seriously, and thus explains why Mary has many fans. After all, there does seem to be something dramatically unique about the epistemic power of the lived aspect of experience. There is something that it is like to see a particular shade of red that is distinctly different than what it is like to see a shade of blue. And it is, at the very least, hard to imagine how we might know what that difference is without having chromatic experiences of our own. A blind person can never know what it is like to see red; a human can never know what it is like to be a bat.

Regardless of whether one accepts Jackson's epistemological intuition, clearly stating the intuition is necessary in order to isolate and identify Jackson's embedded modal intuition. It seems to me that whether the knowledge argument can establish an anti-physicalist conclusion depends primarily on Jackson's modal intuition (it is possible to know all the physical facts achromatically). And if there is reason to reject Jackson's modal intuition, then physicalists should feel free to take any position they want regarding the epistemological intuition.

*Prima facie doubt about Jackson's modal intuition*

Roughly stated, Jackson's modal intuition is that it is possible to know all the physical facts achromatically. On what grounds does Jackson conclude that the facts Mary learns upon leaving the black and white study are not physical? Because, by stipulation, Mary already knew all of the physical facts before she left the black and

white study. Of course Jackson can stipulate whatever he wants to about Mary, but to place her in an achromatic environment and simultaneously stipulate that she knows all the physical facts implies that it is possible to know all the physical facts inside the black and white room. The intuition that Mary knows all the physical facts in an achromatic environment is a modal intuition in that it asserts a possibility: it is possible to know all the physical facts achromatically. Stated so roughly, Jackson's modal claim seems reasonable. Mary could learn everything from black and white televisions, as Jackson suggests. A colleague of mine suggested that, even if just by historical accident, there is nothing in the history of physical science that can only be known by means of chromatic experience.<sup>95</sup>

But the above statement of Jackson's modal intuition is too rough. Clearly, in some sense of 'all the physical facts' - maybe even the first sense that the phrase brings to mind - it is possible to know all the physical facts achromatically. Many of these senses of 'all the physical facts' that would make the modal intuition true can be found in the physicalist literature which charges Jackson with equivocation. Torin Alter argues that, inside the black and white room, Mary does not know all of the facts about physical things, she only knows "all discursively knowable facts"<sup>96</sup>. It is plausible to imagine that all discursively knowable facts can be known achromatically. David Lewis's suggestion that inside the black and white room Mary knows "all the facts which can be taught in lessons" also describes a sense in which Mary's complete achromatic knowledge is possible<sup>97</sup>. The difficulty of justifying Jackson's modal intuition is not finding *some* sense in which all the physical facts can be known in the black and white study, but finding the *right* sense of 'all the physical facts' for the critical connection between the first and



fourth premises of Jackson's knowledge argument. By way of reminder, here is the argument again.

### The Knowledge Argument

1. Inside the black and white study Mary knows *all the physical facts*.
2. Mary learns additional facts when she leaves the black and white study.
3. Thus, there are facts other than *all the physical facts*.
4. If physicalism is true, then there are no facts other than *all the physical facts*.
5. Therefore, physicalism is false.

I would hardly deny that there is some sense of 'all the physical facts' in which premise one is possible. But, and less roughly stated than above, I do deny Jackson's claim that premise one is possible under any interpretation of 'all the physical facts' in which premise four is also true. Instead of making epistemological distinctions to explain what happens when Mary leaves the black and white room, physicalists should adopt the strategy of insisting that the only relevant interpretations of the first premise, the only interpretations in which premise one does its work in the argument, include a sense of 'all the physical facts' in which physicalism asserts that there are no other facts. If it is possible, through purely achromatic means, to learn every fact that can be countenanced by physicalist ontology, then, and only then, does premise one assert a genuine possibility. The question, then, is whether a physicalist worldview can include qualitative facts of the kind referred to in my narrow interpretation of Jackson's epistemological intuition.

Here is a rather odd analogy that captures Jackson's argumentative strategy in the knowledge argument. Think of facts as something akin to little creatures living in a place called Fact Land.<sup>98</sup> One day all of the facts that have physicalist credentials (all those facts which have been certified as consistent with a physicalist worldview) are rounded up and introduced to Mary. Then the local authorities scour the land looking for facts that Mary doesn't know. Such a fact is found when Mary has her first chromatic experience, and the local authorities thereby declare that physicalism is false. Shortly before all of the facts carrying physicalist credentials were rounded up so that they could meet Mary, one particular fact (call it Q) nervously filled out its application for physicalist credentials. Nervously, because Q was not sure what the Department of Physicalist Credentialing would think of a fact such as itself that can only experience the joy of being known by someone who has had chromatic experience. Q thinks of itself as a fact that is as physical as any other fact and defiantly fills out the application for physicalist credentials. On the application Q describes itself as a physical fact about human chromatic experience (specifically about what it is like to see a certain shade of blue) that can only be known by means of chromatic visual states. (As the knowledge argument formally requires, Q is precisely the kind of fact that the epistemological intuition states that you can find in Fact Land.) Of course, Jackson's argument presumes that Q's application will be denied. But, should the Department of Physicalist Credentialing have approved Q's application? It begs the question to respond that Q's application should be denied because Mary doesn't know Q. There needs to be some kind of legislative act in Fact Land establishing criteria by which the Department of Physicalist Credentialing evaluates applications. And remember how high the stakes are; if Q is neither exiled from Fact Land nor granted

physicalist credentials then physicalism is false. *Prima facie*, it seems to me that Q was discriminated against when it wasn't introduced to Mary. Q is a fact about a physical process that is known by means of a biological process. Jackson needs to provide a reason why Q is not physical.

Expressed without reference to Q, the knowledge argument gains traction by making a distinction between those facts that are labeled 'physical' and those that are not. Jackson needs to give us some reason for refusing the label 'physical' to any fact that can only be known chromatically. And the only way to provide such a reason is to offer a definition of 'physical' that is independent of the thought experiment. Put simply: Jackson needs to tell us what makes a fact physical; he can not punt regarding the meaning of 'physical' and still claim that there is any non-question begging reason to believe that it is possible to know all the physical facts achromatically.

### *Jackson defining 'physical'*

But Jackson consistently does punt regarding the meaning of 'physical', even after his 1998 renunciation of the knowledge argument and his conversion to *a priori physicalism*. In his original 1982 presentation of the knowledge argument Jackson begins with "sketchy remarks" that are explicitly not meant to constitute a definition of 'physical'.<sup>99</sup> In a 1986 article written to clarify the knowledge argument and respond to objections, Jackson more or less repeats his sketchy remarks.<sup>100</sup> In a 1994 article defending conceptual analysis, he says that by 'physical' he just means what is usually meant.<sup>101</sup> A 2001 article co-authored by David Chalmers states, "we will not engage the issue of what counts as 'physics'".<sup>102</sup> In 2002, defending the knowledge argument

against common objections before detailing his own reasons for rejecting it, he writes, “exactly how to delineate the physical will not be crucial”.<sup>103</sup> In a forthcoming article, he similarly avoids providing a precise definition of ‘physical’ by writing, “we can side step the issue”.<sup>104</sup>

Without a direct definition of ‘physical’ or ‘physicalism’, there are two strategies defenders of the knowledge argument might take in setting criteria for issuing physicalist credentials, and thus expressing a sense of ‘all the physical facts’ in which premise one is possible and premise four is true. One might follow Jackson’s comments about what Mary *does* know and then argue that the resulting sense of ‘all the physical facts’ is narrow enough to exclude facts like Q while also being broad enough to make premise four true. Alternatively, one might specify the sense in which physicalism is committed to the completeness of the physical facts and then argue that ‘all the physical facts’ in the specified sense excludes facts like Q. In this section I argue that the former strategy fails to exclude Q in any non-question begging way. In the next section I pursue the latter strategy, offering a precise formulation of Jackson’s modal intuition that clarifies the sense of ‘all the physical facts’ in which premise four is true. I then go on to argue that in this latter sense of ‘all the physical facts’ premise one is not possible.

Though Jackson is right to acknowledge that all the physical facts must be taken “very broadly” for premise four to be true, Jackson clearly goes overboard in explaining how broadly ‘all the physical facts’ is to be interpreted when he describes all the physical facts as, “everything which has ever featured in physicalist accounts of mind and consciousness”.<sup>105</sup> Pick your favorite flavor of physicalism (eliminativism, identity-thesis, functionalism), qualitative states are featured in all of these physicalist accounts in

the sense that each provides an account of what to do with Q's application for physicalist credentials. Physicalists such as Alter, Loar, Horgan, and Musacchio would grant Q physicalist credentials; physicalists such as Dennett, Lewis, Conee, and post-1998 Jackson would all have Q exiled from Fact Land.<sup>106</sup> If 'all the physical facts' is interpreted so broadly as to mean "any fact which is true given any position which has ever been advertised as physicalist" then, to answer the knowledge argument, physicalists need only repeat their account of what to do with facts like Q. Jackson is trying to argue that none of these physicalist positions provides an *adequate* account of qualitative states. Yet, in order to make the case that these physicalists accounts are not adequate Jackson needs an independent criteria for determining if a given putative fact is physical.

Jackson's description of all the physical facts as, "everything in *completed* physics, chemistry, and neurobiology, and all there is to know about the causal relational facts consequent upon all this, including of course functional roles" also fails to provide a non-question begging reason to deny physicalist credentials to facts like Q.<sup>107</sup> We can hardly guess what properties and entities might figure in a future natural science. Philosophers at the dawn of the twentieth century might have found it plausible that even a completed physics, chemistry, and biology, could not explain the basic building blocks of life, or how to unlock the destructive power of the atom, or how to (more or less) accurately predict the weather. History suggests that we simply can not place restrictions on the explanatory power of future science from the armchair. And Jackson isn't talking about a *realizable future science*, but an *idealized completed science*: perfect, un-revisable, and without anomaly. To speak metaphorically, Mary is said to be in possession of God's blueprint for the physical universe. Since scientific discoveries

invariably raise new scientific questions, it is not even clear that her storehouse of facts so conceived is finite, in which case she is surely akin to a god herself. But God's intellect is a mystery; we aren't smart enough to imagine the specifics of an omniscient mind. To suggest that anything whatsoever is excluded from an idealized completed science is sheer recklessness at best, and hubris at worst. That Q can currently cite only chromatic experience as a method for being known does not imply that Q could never be known by any possible science; it is plausible to hold that God's blueprint for the physical universe includes Q in a way that no currently imaginable human science could express. Asking whether a particular fact would be a part of completed science is not a practical way of determining whether a particular given fact is physical.

Further, even if it is granted that completed scientific knowledge could not know facts like Q it is unlikely that "completed physical sciences" provides a definition of 'all the physical facts' in which premise four is true. Put simply, physicalism is not committed to the proposition that all facts are scientifically knowable. Consider Colin McGinn's physicalist mysterianism. According to McGinn, the human brain developed from a specific evolutionary history, shaped by the need to survive various challenges in the environment of our ancestors. He also maintains that "our intelligence is wrongly designed for understanding consciousness. Some aspects of nature are suited to our mode of intelligence, and science is the result; but others are not of the right form for our intelligence to get its teeth into, and then mystery is the result".<sup>108</sup> Our brains evolved to perform a certain range of tasks and are thus able to explain only a certain range of things. An explanation of consciousness is outside this range, and thus outside the range of scientific explanation.

I do not endorse McGinn's view. Remember that in Chapter Three I will defend my optimism about a science of consciousness. But observe that McGinn's position has three characteristics: 1) it is coherent to imagine creatures with the kind of limitations McGinn ascribes to human beings. 2) If true, McGinn's position would rule out the possibility of a scientific knowledge of facts like Q. 3) McGinn's position is entirely physicalist; McGinn even offers a physical explanation for why consciousness can not be physically explained. Though McGinn is certainly not a typical physicalist, his position shows that it is possible to be a physicalist without believing that "completed natural science" provides an interpretation of 'all the physical facts'. McGinn presents an empirical possibility, the very existence of which demonstrates that physicalism is compatible with the in principle absence of a scientific explanation of consciousness.

One might object to such use of McGinn, pointing out that all McGinn's position can establish (if true) is that *human* science can not explain consciousness, but that physicalism is nevertheless committed to the existence of an *idealized* scientific explanation. If physicalism is true then there is a possible creature who could provide a scientific explanation of consciousness. Anti-physicalist arguments such as the knowledge argument show, the objection continues, that even a Martian could not scientifically explain consciousness. To see how terribly strange this line of reasoning is, assume that McGinn is right about the impossibility of a human science of consciousness. Further assume that some possible creature could provide scientific knowledge of consciousness. The result would be that no human being could possibly know whether a Martian's claim to have explained consciousness is true. Human beings have no epistemic access to the contents of idealized (as opposed to human) science. If it is

granted that human science can not explain consciousness, the result is that it has also been granted that there is no available evidence that idealized science could nonetheless explain consciousness. Especially since so many as yet unimagined facts will turn out to deserve physicalist credentials, there is no criteria by which one could confidently decide that facts like Q ought to be excluded.

*All the truths about a minimal physical duplicate world*

Beginning in 1994 the focus of Jackson's work became, roughly, to defend the view that physicalism is only true if there are no non-physical facts, which is premise four of my the knowledge argument ("If physicalism is true then there are no facts other than all the physical facts"). Before 1998 Jackson believed there were facts other than all the physical facts and concluded, by *modus tollens*, that physicalism is false. After 1998 Jackson believes that physicalism is true and concludes, by *modus ponens*, that there are no facts other than all the physical facts. In either case, Jackson's philosophical writing since 1994 has increasingly focused on capturing the sense in which physicalism claims that the physical facts are complete.

In describing the sense in which physicalism is committed to the completeness of the physical facts Jackson appeals to the following supervenience thesis: "any world which is a *minimal* physical duplicate of our world is a duplicate *simpliciter* of our world".<sup>109</sup> The use of 'minimal' is essentially a stop clause in constructing a world; any world that is a physical duplicate of ours *and contains nothing else* is a duplicate *simpliciter* of our world. As Jackson expresses it, "a minimal physical duplicate of our world is what you would get if you – or God, as it is sometimes put – used the physical



nature of our world (including of course its physical laws) as a recipe in this sense for making a world".<sup>110</sup> A possible world that is a physical duplicate of our world and also contains ectoplasm is not a duplicate *simpliciter* of our world, but such a world is clearly no threat to physicalism. In essence, the use of 'minimal' here acknowledges that physicalists do not need to hold that physicalism is true in every possible world, only in the actual world.<sup>111</sup>

Jackson's argument that physicalism is committed to this supervenience thesis is straightforward and convincing. First, if the minimal supervenience thesis is false then physicalism is false. There would be something in our world that is not contained in a minimal physical duplicate of our world; it would follow that there is something non-physical in our world and physicalism would be false in that it would be incomplete. Second, if physicalism is false then the minimal supervenience thesis is false: if physicalism is false then our world contains some non-physical nature that is not contained in a minimal physical duplicate world, and thus such a world is not a duplicate *simpliciter* of our world. Therefore, the minimal supervenience thesis is false (or true) if and only if physicalism is false (or true). I agree with Jackson that the minimal supervenience thesis is a genuine commitment of physicalism and that it is perhaps the best way of capturing the sense in which physicalism is committed to the completeness of the physical facts. Thus, we have found a true interpretation of premise four of Jackson's knowledge argument that captures Jackson's intention: "If physicalism is true then there are no facts other than all the facts that obtain in a minimal physical duplicate of our world". 'All the physical facts' are all the facts that obtain in a minimal physical duplicate. The Department of Physicalist Credentialing should accept the application of

any fact, and only those facts, for which they find a doppelganger in their newly constructed minimal physical duplicate world.

Recall that Jackson's argument relies on there being a sense of 'all the physical facts' in which premise four is true and premise one is possible. The knowledge argument can only establish its anti-physicalist conclusion if Mary knows all the physical facts in the very same sense of 'all the physical facts' in which physicalism claims that there are no facts other than all the physical facts. Having found the sense of 'all the physical facts' in which premise four is true, it is now possible to present a precise formulation of Jackson's modal intuition.

Jackson's Modal Intuition: It is logically possible to know all of the facts about a minimal physical duplicate world (including all of the facts about chromatic experience in such a world) without having chromatic experience.

Returning to the Fact Land analogy, Jackson's modal intuition is that Q should be denied physicalist credentials because Q would not be true in a minimal physical duplicate world. To avoid over-reliance on a single thought experiment, I will soon say nothing more about our friend Q even though I still think Q has been discriminated against. As Q herself might emphatically point out, Q is a fact about chromatic experience and chromatic experience is a physical process of certain biological organisms. To refuse Q physicalist credentials merely because she can only be known by means of chromatic experience seems like discrimination. In a minimal physical duplicate world there are the same physical processes of replica biological organisms. Q has a doppelganger in a

minimal physical duplicate world because in a minimal physical duplicate world there are replicas of each organism that exists in our world, whose brains are in the exact same states and are thereby having the same experiences. Since Q is a fact about such experiences, Q also exists in a minimal physical duplicate world and is thereby entitled to physicalist credentials on the only sense of ‘all the physical facts’ available that fits Jackson’s argumentative needs. In part III, I make a more careful argument rejecting Jackson’s modal intuition by showing the similarity between Chalmers’ and Jackson’s modal intuitions. Before doing so it will be helpful to restate the knowledge argument in terms of the distinction between the epistemological and modal intuitions.

*Jackson’s two intuitions: the knowledge argument restated*

Whereas Jackson acknowledges that the knowledge argument rests on a single intuitive premise (no amount of physical information can tell you what it is like to, for instance, see red), I find two intuitive premises.

Jackson’s Epistemological Intuition: chromatic experience is the only possible method for learning certain facts about the qualitative properties of chromatic experience.

Jackson’s Modal Intuition: It is logically possible to know all of the physical facts about a minimal physical duplicate world without having chromatic experience.

The striking image of Mary leaving the black and white study and seeing a red rose for the first time provides the epistemological intuition. The stipulation that she knew all the physical facts (in the sense relevant to the truth of physicalism) before she left the black and white study asserts the modal intuition. Anyone who is willing to accept both intuitions is forced to conclude that our world includes some facts that are not part of a minimal physical duplicate world, and ultimately that physicalism provides an incomplete picture of our world. I endorse Jackson's epistemological intuition and reject his modal intuition. Unless one has had chromatic experiences of one's own, one can not know that when other people have a red experience they are having *that* experience (the qualitative experience we call "seeing red"). However, in a minimal physical duplicate world when people have a red experience they are also having *that* experience, and thus knowing all the facts that obtain in a minimal physical duplicate of our world requires knowing what experience they are having. In the next section, I argue that Jackson asserts a logical impossibility when he says, "Inside the black and white study Mary knows all the physical facts". I claim that knowing all the physical facts requires chromatic experience.

### III. Rejecting Jackson's Modal Intuition

I will argue that that Jackson's modal intuition is ultimately the same kind of modal intuition behind Chalmers' zombie argument. Chalmers' zombie intuition is routinely rejected by physicalists, and with good reason. My strategy is to first argue that Jackson's and Chalmers' modal intuitions are closely related before arguing that such modal intuitions ought to be rejected.

*'Chromatic experience' and 'logically possible to know'*

To see the similarity between Jackson's and Chalmers' modal intuitions, it is helpful to notice two things about Jackson's thought experiment.

Uncontroversially, Jackson is using color merely as an example of qualitative experience, an example that is both particularly striking in itself and convenient for setting up a clear thought experiment. The knowledge argument could have been expressed through any number of related thought experiments: instead of a black and white room Mary might have lived in a sound proof room and known all the physical facts about auditory experience; she might have been fed entirely by gastronomy tube and known all the physical facts about taste experiences, and so on. The point is that qualitative experiences have properties that are, plausibly, not captured by a physicalist worldview. If the knowledge argument works as an argument about chromatic experience it is hard to see why it would not also work as an argument about any particular qualitative experience and thereby about qualitative experiences generally. The broad nature of Jackson's intuitions should not be overlooked simply because Jackson uses chromatic experience as his example.

Only slightly controversially, Mary is meant to be an idealized knower who knows everything that is *knowable in principle*. If human intelligence is just not up to the task of knowing all the physical facts then Mary must have super-human intelligence. If there are physical facts that can only be known by means of some as not yet invented technological gizmo, then be sure that Mary has such a gizmo. Since Mary knows everything that can in principle be known about the physical properties of our world, any

fact which is outside her knowledge demonstrates the incompleteness of physicalism *by its very existence*. Therefore, and here is the part that might be controversial, Jackson's use of Mary as an ideal knower (knowing every physical fact which there is) is an epistemic route to the property relations themselves. As Jackson sees it, we can access the metaphysical issues about what physical properties there are by considering Mary, the idealized knower of the physical facts. If *something can't be known by Mary*, then it's *not a physical fact*; her pre-release knowledge exhaustively quantifies over all physical properties. If there are properties not quantified over by her knowledge, such properties are not physical; that is, any properties not quantified over by Mary's pre-release knowledge are not part of a minimal physical duplicate world.

Taking the above two observations together we can see that Jackson's modal intuition is a species of the following modal claim: there is a possible world that is a physical duplicate of our world but contains no qualitative experience. If such a world is possible, then Mary can know all the truths that obtain in a minimal physical duplicate world without knowing about qualitative experience, and without having qualitative experiences herself. If such a world is possible, premise one of the knowledge argument asserts a genuine possibility and, if Jackson's epistemological intuition is sound (as I think it is), the knowledge argument establishes its anti-physicalist conclusion. If no such world is possible, then any world that is a physical duplicate of our world also contains facts that can only be known by means of qualitative experience. If a minimal physical duplicate world contains facts that can only be known on the basis of chromatic experience, then premise one of the knowledge argument does not assert a genuine

possibility, and Jackson's epistemological intuition is irrelevant to the truth of physicalism.

### *Zombie University*

I hope to prove that David Chalmers' zombie intuition is a species of the same modal intuition as Jackson's Mary intuition. Chalmers describes his own zombie twin as a physical and functional replica of himself. It is a physical replica in that for every physical particle in Chalmers' body there is a corresponding identical particle in Chalmers' zombie twin. It is also a functional replica in that, for instance, Chalmers' zombie twin would give the same answer to any question that the real Chalmers would. The only difference between Chalmers and his zombie twin is that, unlike the real Chalmers, the zombie twin has no qualitative experience; there is nothing that it is like to be a zombie. At the global level, Zombie World is a physical and functional replica of our world in which qualitative experience is entirely lacking. There is nothing that it is like for anyone to live in Zombie World. Chalmers' modal intuition is that zombie twins and Zombie Worlds are logically possible. According to Chalmers if zombies are logically possible then the facts about qualitative experience do not logically supervene on the physical facts. If there is a possible world where all of the physical facts are the same as in our world, but where the qualitative facts are different than in our world, it follows that there is no necessary relationship between physical and qualitative facts in our world.

Though Chalmers does not use the language of 'minimal physical duplicate' Jackson's minimal supervenience thesis is clearly one way to capture his intention. After all, a minimal physical duplicate world is what you get when you construct a world that is

a physical and functional replica of our world.<sup>112</sup> And so an assertion that Zombie World is logically possible is equivalent to the assertion that it is logically possible to know all of the physical facts without knowing the qualitative facts.<sup>113</sup> In both cases the claim is ultimately that the physical facts about our world do not entail the qualitative facts; physical facts and qualitative facts can independently vary across worlds.

In fact, Chalmers' anti-physicalist argumentative strategy closely parallels the understanding of the knowledge argument that emerged when Jackson's two intuitions were distinguished from one another. Chalmers begins *The Conscious Mind* by arguing that consciousness needs to be taken seriously, that qualitative experiences are strikingly real events: there is something that it is like to see red, to feel pain, to hear music, and to be angry.<sup>114</sup> As noted earlier, Chalmers' call to take consciousness seriously is equivalent to my narrow interpretation of Jackson's epistemological intuition. Jackson adds the clarification that qualitative experience is to be taken seriously both as a method for learning certain facts as well as a topic for study. Paralleling Jackson's modal intuition, Chalmers next move is use a variety of thought experiments (zombies, Mary, and inverted spectra) to establish the failure of the qualitative facts to logically supervene on physical facts. Given that the failure of the qualitative facts to logically supervene on the physical facts establishes the falsity of reductive physicalism, Chalmers then argues that a naturalistic form of property dualism is the most prudent metaphysical position remaining.

Consider Chalmers' own assessment of the knowledge argument,



This argument [the knowledge argument] is closely related to the arguments from zombies or inverted spectra, in that both revolve around the failure of phenomenal facts to be entailed by physical facts. In a way, they are flip sides of the same argument. As a direct argument against materialism, however, Jackson's argument is often seen as vulnerable due to its use of the intensional notion of knowledge. Many attacks on the argument have centered on this intensionality – arguing, for example, that the same fact can be known in two different ways. These attacks fail, I think, but the most straightforward way to see this is to proceed directly to the failure of supervenience, which is cast in metaphysics rather than epistemology.<sup>115</sup>

This passage not only shows that Chalmers agrees that his zombie intuition revolves around the same issue as the Mary intuition but also that, and more importantly, Chalmers points out that the best defense against the charge that the knowledge argument equivocates on 'knowledge' is to "proceed directly" to strictly metaphysical questions about supervenience. In other words, Chalmers seems to indicate that the primary difference between his own anti-physicalist arguments and Jackson's is that the knowledge argument is strategically weakened by the focus on what it is possible *to know*. On Chalmers' reading, the knowledge argument would be strengthened and the core intuition (the failure of the qualitative facts to supervene on the physical facts) would be made clearer by avoiding a discussion of knowledge. My argument that Jackson can describe the completeness of Mary's knowledge in no non-question begging way provides argumentative support for Chalmers claim that Jackson is better off initially

raising the issue of supervenience. Indeed, I proceeded in a similar way when I turned to premise four and sought a sense of 'all the physical facts' in which premise four is true. Chalmers view is that the knowledge argument ultimately works, but the focus on knowledge (instead of supervenience) is a distraction.

Perhaps the most direct way to see the relationship between Jackson's Mary intuition and Chalmers' zombie intuition is to notice that if either is true (or false) then the other must be true (or false). If Jackson's modal intuition is correct then not only is Zombie World logically possible, Mary could tell you everything you need to know about how to build it. After all, Mary knows everything there is to know about a minimal physical duplicate world and that is exactly what the Zombie World is. If the zombie intuition is true then it is possible to know all of the facts about a minimal physical duplicate world without knowing the qualitative facts. After all, the zombie intuition asserts that the qualitative facts that obtain in our world do not obtain in a minimal physical duplicate world. Even if Jackson's modal intuition and Chalmers' zombie intuition are not ultimately expressions of precisely the same modal intuition they at least have the same truth value.

Mixing the two thought experiments offers a colorful way to demonstrate the relationship between them. We might think of the black and white room as a kind of Zombie University where Mary is the star student. After all, (assuming that Jackson's modal intuition is correct) you could learn everything that there is to know about the Zombie World inside the black and white study. Of course, zombies are welcomed to attend Zombie University (their lack of qualitative states being no barrier to learning what is taught inside the black and white room). Nevertheless, the main point of putting

the university inside the black and white room is to mimic a zombie-like existence for human students such as Mary. Mary is easily the star student, as she is the only one ever to learn everything that the university has to teach. Jackson's epistemological intuition can even be recast as a critique of the school: there are things they don't teach at Zombie University.

Both the zombie intuition and Jackson's modal intuition are creative ways of asserting the failure of qualitative facts to logically supervene on physical facts. Therefore, if there are grounds to believe that qualitative facts logically supervene on the physical facts then there are grounds for rejecting the anti-physicalist arguments of both Jackson and Chalmers. In what follows I attempt to provide such grounds, moving back and forth between Mary and zombies as convenience of locution dictates.

Next I will argue that Jackson's and Chalmers' modal intuitions are unsound. Even if – as I anticipate – the following argument fails to persuade any philosophers who accept Jackson's modal intuition (including physicalists such as Jackson post-1998, who still accepts the modal intuition embedded in the knowledge argument), the preceding arguments demonstrate that the knowledge argument fails to achieve Jackson's original stated goal. In Jackson's original presentation of the knowledge argument, Jackson contrasted the knowledge argument with the modal argument (akin to Chalmers' zombie argument) and indicated that, unlike the modal argument, the knowledge argument does not ultimately rest on a disputed intuition. Jackson contends that the *epistemological* intuition at the heart of the knowledge argument is both more plausible and more widely accepted than the *modal* intuition at the heart of the modal argument. If the knowledge argument itself rests on a similar modal intuition (independent of the argument's

plausible and widely accepted epistemological intuition) then the knowledge argument is, in the end, no improvement on the anti-physicalist arguments that were already available. If anything, the knowledge argument turns out to be weaker than the modal argument in that it carries the additional epistemological burden. After all, the primary difference between Chalmers' zombie intuition and Jackson's intuition about the possibility of a physically omniscient achromatic knower is Jackson's focus on what is known. The irony is that the knowledge argument has seemed attractive to people who have been turned off by Chalmers' more naked appeal to the disputed intuition; however, Jackson ultimately makes a clothed appeal to the same intuition.

### *Contradictions and mis-descriptions*

Recognizing that many do not share his intuitions about zombies, Chalmers accurately describes the landscape for arguing about modal intuitions. Chalmers acknowledges that argumentation about modal intuitions generally amounts to little more than tub-thumping proclamations, but he says that the logical possibility of zombies seems perfectly obvious to him. According to Chalmers, since a zombie can be coherently described, the burden of proof falls on those who deny the logical possibility of zombies. Those wishing to deny the intuition, he says, "must give us some idea of where a contradiction lies, whether implicit or explicit"<sup>116</sup>; they must show that the description of a zombie involves some kind of mis-description.<sup>117</sup> I intend both to 1) give an idea of where the contradiction lies, and 2) characterize the mis-description that is involved with the modal intuition.

The implicit contradiction involved in imagining physically omniscient achromatic knower such as Mary can be stated very simply: Jackson explicitly states that Mary does and then implicitly states that she does not know all the physical facts about chromatic experience. He explicitly states that she does know all the physical facts; but by locking her in a black and white study it seems to me that he has implicitly said she does not know all the physical facts. Jackson wants us to interpret ‘fact’ broadly enough to cover the kind of epistemic event pointed to in the epistemological intuition. I grant the epistemological intuition; there are such facts. The issue relevant to the soundness of the modal intuition is whether such facts would also be true in a minimal physical duplicate world; it seems perfectly obvious to me that they would. One can think about the issue in terms of the physical sciences. The scientific facts about brains and the like would certainly be true in a minimal physical duplicate world. Therefore, my counterpart in a minimal physical duplicate world has the exact same brain structure and fundamental particles that I do; he has the exact same neural inputs that I am currently having; and he is making the same decisions that I am regarding those inputs. Currently, my counterpart in a minimal physical duplicate world is sitting outside writing. He turns his head in the direction of the yard and visual input at some specific wavelength hits his pupils, is converted into an electrical-chemical message, and is then sent to his occipital lobes via his optic nerve. In the occipital lobe my minimal-physical-world doppelganger creates a visual orientation map of the yard and information corresponding to the specific wavelength of the light is sent to his V4 cells and his visual orientation map is filled with the color he calls “green”. And thus it is true of my minimal physical world doppelganger that he is having *that* experience (the experience he calls “seeing green”). And if Mary

doesn't know the fact that my doppelganger is having *that* experience when he looks at green, then she doesn't know all the facts about my minimal physical doppelganger. How we are able to explain and understand such qualitative facts about me and my doppelganger is an important, but separate, issue. In order to defend physicalism from the knowledge argument, however, all that is necessary is to claim that whatever you say about my experience when I see green (learning new facts, acquiring new abilities, learning phenomenal concepts), the same should also be said of a minimal physical duplicate of me. And similarly for Chalmers' zombie intuition. A perfect physical replica of me is in the same brain state, with the same neural impulses, the same emotional state etc., and is necessarily also having the same qualitative experiences that I am having.

Having located what I take to be the implicit contradiction embedded in the description of the Mary and zombie thought experiments, I turn to the sense in which I consider the apparently coherent descriptions of Mary and zombies to involve a mis-description. Jackson's description of Mary's complete physical knowledge conflates metaphysical and epistemological issues. For the knowledge argument to work, Mary's knowledge must be complete in the metaphysical sense required by premise four of the knowledge argument, and yet Jackson's descriptions of Mary's knowledge invariably involve reference to various epistemological strategies. My argument that such epistemological strategies ("completed physical science") do not capture the relevant sense of completeness can be found above. To understand the subtlety of the problem it is helpful to see the sense in which Chalmers' description of zombies involves a version of the same mis-description, especially as Chalmers' advertises the zombie intuition as epistemology-free in comparison with the knowledge argument.

The parallel in Chalmers' argument to the sense in which Mary is said to have complete physical knowledge is the sense in which my zombie twin is said to be a complete physical and functional replica of me. Though Chalmers starts out his description of how to construct a zombie twin purely in the metaphysical terms of replicating all of the physical properties and functional properties, when he turns to explaining how it is possible that a physical and functional replica might lack consciousness, he turns to epistemological considerations about what can be entailed on the basis of various physically privileged epistemological strategies. In short, Chalmers offers a bait and switch: describing first a zombie who is a metaphysical replica (my zombie twin is composed of the same arrangement of physical particles), and then offering an argument that an epistemological replica (a zombie twin who is not distinguishable from me by means of the physical sciences) might lack qualitative experience. In his account of why zombies are possible, Chalmers points out that the facts about microphysics do not entail propositions about consciousness, and neither do the facts about functional psychology. In other words, he points to the same ways of knowing physical facts (physical sciences, functional psychology) that Jackson points to when describing what Mary does know in the black and white study. Earlier I argued that Jackson's account of Mary's complete physical knowledge ("everything ever featured in physicalist accounts", "completed physical science") fails to exclude the kind of qualitative knowledge that Jackson's epistemological intuition points to in any non-question begging way. To the same extent, Chalmers fails to provide a non-question begging reason for excluding such properties when constructing a zombie. That

qualitative properties can not be known by means of the physical sciences does not mean that qualitative properties are properties above and beyond physical properties.

The modal intuitions of both Jackson and Chalmers rest on a mis-description of the sense in which physicalism claims completeness. As I see it, physicalism is committed to completeness only in the sense of the minimal supervenience thesis. That is, once the facts that are undisputedly physical facts (biological and functional facts, for instance) were fixed there was no more work for God to do. It is a mis-description of physicalism's claim to completeness to suggest that the physical sciences must be suitable for knowing everything. Granting for the sake of argument (and only for the sake of argument) that there are facts about the world that can not be explained by means of physics, chemistry, biology, and functional psychology; nonetheless such facts would necessarily be true given the entire set of the kind of facts that can be known by means of the physical sciences and functional psychology. If God were to construct a complete duplicate of our world all she would have to do is to place each physical particle, atom, molecule, cell etc. in the same relation to one another as they are in our world and she would be done.

## Conclusion

When discussing McGinn earlier, I indicated that I am optimistic about a science of qualitative experience; my optimism is thus far undefended. In this chapter, I have argued that progress sorting out the relative roles of epistemological and metaphysical considerations in the knowledge argument sheds light on why the argument fails to demonstrate the incompleteness of physicalist ontology. In the next chapter, I will make a



parallel argument that such progress also sheds light on why the knowledge argument fails to demonstrate the incompleteness of scientific explanation.

---

Notes to Chapter Two

- 91 That is, I am not merely arguing that her visual cortex would decay (though surely it would).
- 92 Jackson, "Mind and Illusion," 422.
- 93 Horgan, "Jackson on Physical Information and Qualia."
- 94 See the Chapter One discussion of David Lewis in particular.
- 95 Thanks to Ronald Loeffler for pointing out this objection, and for reviewing a much earlier draft of this chapter.
- 96 Alter, "A Limited Defense of the Knowledge Argument," 47.
- 97 Lewis, "What Experience Teaches."
- 98 I do not advise thinking of facts as little entities in this way. Nevertheless, Jackson's account of facts in the knowledge argument lead us to think of facts in exactly this strange kind of way.
- 99 Jackson, "Epiphenomenal Qualia," 39.
- 100 Jackson, "What Mary Didn't Know."
- 101 Jackson, "Armchair Metaphysics."
- 102 Jackson and Chalmers, "Conceptual Analysis and Reductive Explanation," 316.
- 103 Jackson, "Mind and Illusion," 421.
- 104 Jackson, "A Priori Physicalism."
- 105 Jackson, "Epiphenomenal Qualia," 42.
- 106 Which is not to say that Lewis and Conee argue for eliminativism. They would have Q exiled from Fact Land in that Q is not a fact; it might be said that they would have Q deported to Ability Land.
- 107 Jackson, "What Mary Didn't Know," 51.
- 108 McGinn, *Mysterious Flame*, xi.
- 109 Jackson, "Armchair Metaphysics," 28. Jackson's interpretation of this supervenience thesis is quite at odds with my use of it in this chapter. For a description and critique of Jackson's claims regarding the implications of this supervenience thesis, see Chapter Four.
- 110 Ibid, 28.
- 111 Jackson's formulation rules out trivial ways to make physicalism false. Since most philosophers think it is metaphysically possible for there to be non-physical things (though not all of course) any definition of physicalism that merely says that two physically duplicate worlds are duplicates simpliciter is inadequate. A possible world which is a physical duplicate of ours and contains extra non-physical stuff is no threat to physicalism.
- 112 Actually, 'functional' here is not necessary. Chalmers includes it merely to demonstrate that he concedes that functional descriptions are ultimately physical. The point is that even functional descriptions can not capture what it is like to be in a particular qualitative state.
- 113 That is, these two claims are equivalent unless one holds a view like McGinn's that not all physical facts can be known by human beings.
- 114 Chalmers, *Conscious Mind* 4 - 11.
- 115 Ibid, 140.
- 116 Ibid, 99.
- 117 Ibid, 97.

## Chapter Three:

### Qualitative Facts and Scientific Epistemology

In Chapter Two I argued that taking consciousness seriously does not require abandoning physicalism. In Chapter Three I will argue that taking consciousness seriously does not require abandoning scientific explanation. Qualitative facts are real, physical, and explainable.

#### I. Accepting the Epistemological Part of the Knowledge Argument

Even some physicalist philosophers claim that Jackson's knowledge argument establish that science can not explain qualitative facts. For example, Colin McGinn cites the knowledge argument to support his position that consciousness is a fundamental mystery (though he also believes that consciousness is ultimately physical).<sup>118</sup> Joseph Levine argues that the thought experiment about Mary establishes the existence of an epistemological gap, but not a metaphysical gap.<sup>119</sup> That is, Levine argues that Jackson's thought experiment demonstrates the incompleteness of physicalist explanation but not the incompleteness of physicalist ontology. Jackson's knowledge argument is commonly cited by non-reductive physicalists as evidence for their rejection of reductive approaches to explaining consciousness. All such approaches to the knowledge argument can be summarized as accepting the epistemological claim of Jackson's argument while denying the metaphysical claim.

The response to the knowledge argument that I offered in chapter Two also accepts Jackson's epistemological intuition while denying his metaphysical (specifically,

his modal) intuition. I see the relationship between metaphysics and epistemology in the knowledge argument differently than McGinn, Levine, or Jackson. By way of reminder, in the Introduction, I offered the following formulation of the knowledge argument:

#### The Knowledge Argument

1. Inside the black and white room Mary knows all the physical facts.
2. After her release from the black and white room Mary knows new facts.
3. Thus, there are facts other than all the physical facts.
4. If physicalism is true then there are no facts other than all the physical facts.
5. Therefore, physicalism is false.

As the argument is commonly understood, the first sub-argument (lines one through three) makes an epistemological claim; the second sub-argument (lines one through three) makes a metaphysical claim. By contrast, in chapter Two I argued that the first premise implies the modal claim that it is possible to learn all the physical facts in an achromatic environment. While I acknowledge that it is possible to learn all the physical facts inside a black and white room in some sense of ‘all the physical facts’, I argued that it is not possible to learn all of the physical facts in a black and white room in the sense of ‘all the physical facts’ in which premise four is true. Thus, I reject Jackson’s implied modal intuition that it is possible to know all of the facts countenanced by physicalist ontology without qualitative (specifically, chromatic) experience. Because of the crucial link between the first premise and the fourth premise, the first sub-argument involves both an epistemological and a metaphysical (specifically, modal) claim.<sup>120</sup> Thus, contrary to Jackson’s account of the epistemological intuition involved in the knowledge argument

(“no amount of physical information can tell you what is like to see red”) I offered the following much narrower formulation of Jackson’s epistemological intuition.

Jackson’s Epistemological Intuition: qualitative experience is the only possible method for learning certain facts about qualitative experience.<sup>121</sup>

The reason it seems “just obvious” to Jackson that Mary learns something when she leaves the black and white study is that Mary’s environment in the study was achromatic. That Mary knew all of the physical facts is a stipulation of the thought experiment that is above and beyond the core epistemological intuition. I contrasted Mary with Harry, an otherwise ordinary bloke who never leaves the black and white room. Asserting that there are things Harry never knows captures Jackson’s epistemological intuition just as well as asserting that Mary learns something new when she leaves. Stating Jackson’s core epistemological intuition requires reference no ‘physical’ or ‘physicalism’. There is a strong intuition that Mary learns something when she leaves the black and white study because there are facts about what it is like to have chromatic experience that you can not know unless you have had chromatic experiences of your own.

For the sake of clarity, I reserve the term ‘qualitative facts’ for the facts that Jackson’s epistemological intuition refers to. By rough approximation, we might say that qualitative facts are most naturally expressed with propositions such as ‘it is like *that* when someone sees red’. There is something that it is like to see red, feel pain, taste Vegemite, or listen to Bach; qualitative facts are the facts about what it is like to have such experiences.

In his emphasis on first-person epistemic access, Jackson is following in the footsteps of Thomas Nagel's central argument in his landmark article, "What is it Like to be a Bat?".<sup>122</sup> In this article Nagel's point is ultimately a simple one: you can not know what it is like to be a bat unless you are a creature that is like a bat. Our knowledge of what it is like for other human beings to, for instance, see red is based on analogy from our own case. Since we can not analogize from our own case to the case of what it is like to use sonar, we can never know what it is like to be a bat. Yet, there is something that it is like to be a bat; there is some property of bat experience that is qualitative in nature – sometimes called the "Nagel-property". Nagel's epistemological intuition is that the Nagel-property can only be known from a first-person point of view. And so it is with Mary: before leaving the black and white study Mary could not have known what it is like to have a chromatic experience because she had not yet had one. Put most generally, first-person qualitative experiences are uniquely informative.

Whether there are qualitative facts is one issue; whether such facts are consistent with physicalism, or are scientifically knowable, are separate issues. In Chapter Two, I pointed out that Jackson's stipulations about Mary already assume that qualitative facts are not physical facts in their own right. Similarly, the very notion of a scientifically omniscient black and white knower merely assumes a complete scientific knowledge can be had without qualitative (specifically chromatic) experience. When asking if Jackson's epistemological intuition is inconsistent with a science of consciousness it is less than helpful to rely on thought experiments that merely presume that the qualitative facts are over and above the complete scientific facts.

Nevertheless, it will be helpful to have some kind of thought experiment to work with. Instead of Mary, imagine Mary's twin sister Hermy, the subject of a thought experiment paralleling the thought experiment about Mary while also avoiding the question begging assumption involved in presuming epistemic completeness in the absence of qualitative experience. Such thought experiments are the best way to determine if Jackson's epistemological intuition supports the conclusion that qualitative facts can not be scientific. That is, if there can be no science of qualitative facts, Hermy ought to be able to show us why.

Hermy has acquired a complete scientific knowledge; she knows everything that actual human science could possibly discover. Pick any field of science, she knows it all. Further, grant for the sake of argument, that a complete scientific knowledge will turn out to be a whole lot more (but not infinitely more) of the same sort of stuff that we have come to expect from science. Hermy does not possess a divine intellect, she's just *really* smart (much more so than you or I) and does science for a *really* long time (much longer than an ordinary human lifetime).<sup>123</sup> Jackson's knowledge argument begs the question by denying Mary color experiences; Hermy is free to investigate by whatever means she sees fit. (Fortunately, she does not know about the oppressive conditions that her twin sister is forced to live in.) Hermy has a limitless scientific curiosity and is intellectually satisfied with the results of all her experiments. She might be described as the embodiment of the entire community of scientists, or as the idealized (human) scientific knower.

Hermy has just entered graduate school and her advisors have set her the task of writing a dissertation titled, "The Correct Scientific Explanation of What it is Like to See

Red”. Hermy is completely convinced that Jackson’s epistemological intuition is correct. As Hermy sees it, there are qualitative facts. In fact, her advisors have told her that her dissertation will be rejected if she denies Jackson’s epistemological intuition. The question is: will she ever finish her dissertation? It might seem “just obvious” that she will not. Hermy can describe in elaborate detail the *correlations* between the phenomenal experience of seeing red and activity of V4, and she makes important discoveries. Yet, no matter how much she studies the activity of V4 cells she is unable to explain why all the electrical-chemical activity in that region of the brain gives rise to the particular experience seeing red. Having read David Chalmers in her spare time, she still finds it conceivable that all of this buzzing of V4 might go on without any phenomenal red experience. She has thousands of pages of draft material, but it all seems to leave out the “redness of seeing red”.<sup>124</sup>

McGinn, Levine, and many others would claim that Hermy is doomed to become the clichéd graduate student who steadfastly takes on an impossible project spends the rest of her life in graduate school. If Hermy (the ideal human scientific knower) can’t provide a scientific explanation of what it is like to see red, then there is no such scientific explanation to be had. I contend that there is no *a priori* reason for concluding that Hermy will never finish her dissertation. On the contrary, trends in the philosophy of science indicate reason for optimism.

## II. Jackson’s Epistemological Intuition and the Explanatory Gap

One of two approaches might be taken in arguing that Hermy will not be able to finish her dissertation. First, one might assert that it is not possible to know what the

qualitative facts are in any scientifically credible way because scientific facts are objective facts and qualitative facts are subjective. Colin McGinn, for example, believes that the human brain has evolved the cognitive capacities to explain certain kinds of things but not others. Though McGinn frames his position in terms of a failure of scientific explanation of qualitative facts, the only reason McGinn provides for the conclusion that science can not explain qualitative states is that qualitative states can only be known from the first-person point of view. And no amount of first-person reporting amounts to third-person scientific knowledge.<sup>125</sup> There is a certain something about lived experience that can not be captured by propositions accessible from a purely third-person point of view. Thus, although McGinn frames his position in terms of the inability of science to explain qualitative facts, the problem, as McGinn sees it, is that the qualitative facts can not be known in any scientifically credible way in the first place. We know that there are qualitative facts, and we know roughly what they are, but such knowledge is not scientific. Hermy could never finish her dissertation because qualitative properties are just not the kind of properties that science can access.

Second, one might assert that there can be no scientific explanation of *why* the qualitative facts are as they are, because propositions expressing qualitative facts don't connect in the right way with propositions that express the relevant physical, chemical, or biological facts (that is, propositions of the traditional natural sciences). Joseph Levine (and many others) argues along these lines, primarily on the basis of absent and inverted qualia thought experiments. As Levine puts it, "no matter how rich the information processing or the neurophysiological story gets, it still seems quite coherent to imagine that all that should be going on without there being anything it's like to undergo the states



in question”.<sup>126</sup> Since no amount of neurophysiological detail entails either that there are qualitative states (thus allowing for absent qualia) or that the qualitative facts are as they are (thus allowing for inverted qualia), the traditional natural sciences can not explain qualitative facts. Even if the qualitative facts are discoverable in some reliable way, the natural sciences will not be able explain them.

Though science certainly does not operate by the neat and clean process of first figuring out *that* something is the case before explaining *why* it is the case, the *that-why* distinction usefully organizes the discussion by separating the relevant issues. In the next two sections I argue that qualitative facts may be scientifically knowable in their own right and that qualitative facts can be scientifically explained. Though I will unavoidably engage in much speculation about what a science of consciousness might look like, I am not trying to determine either the appropriate methodology or the appropriate theoretical model of such a science. Ultimately, I am optimistic about a science of consciousness. My current goal is merely to argue that Jackson’s epistemological intuition need not temper optimism about the development of a science of consciousness. When interpreted narrowly enough, Jackson’s epistemological intuition is not an insurmountable barrier to a science of consciousness, but a valuable resource for getting such a science off the ground.

### III. Knowing *That*: First-Person Epistemic Access and the Demarcation Problem

The central claim to Jackson’s epistemological intuition is that you can only know qualitative facts if you have the qualitative experience yourself. While there is no consensus regarding the precise criteria for distinguishing scientific from unscientific

knowledge, or even agreement that there are precise criteria, few if any of the proposed criteria rule out the kind of first-person knowledge that Jackson's thought experiment directs one's attention to. Karl Popper's *criterion of falsifiability* does not rule out first-person phenomenological reports.<sup>127</sup> Claims about qualitative facts can be falsified by those who are willing to check them against their own experience. Claims about qualitative facts are even objective in the sense relevant to demarcation. Sure, they are reports about subjective experience, but there is no reason to assume that they will be tainted by the particular experiences of either Hermy or her research subjects. Bias reports can be detected by comparing them with other reports, just as bias claims in physics are detected when other physicists are unable to duplicate results. Perhaps qualitative facts are not expressible in discursive language, and can only be indicated demonstratively. This too does not prevent them from being scientifically known; there would just have to be some mechanism of pointing in any account of them. Maybe Hermy's dissertation will include swatches of various shades of red sprinkled throughout, along with claims such as "red swatch 43 is such and such". That might prevent certain people (such as the color blind) from precisely understanding the claim, but something can be true without being available for universal understanding.

Such first-person knowledge is certainly unusual in scientific discourse, and far from the ideal of data that is based on direct third-person observation. But data based on direct third-person observation is the ideal for pragmatic reasons, not reasons having to do with demarcation. Direct third person observation is ideal because such data is clean, clear, and almost impossible to deny. Yet, research that is uncontroversially scientific often fails to meet this ideal. Statements about quarks do not meet the idealized standard,

neither do statements about the ancestral connection between horses and zebras, neither do any statements regarding what happens in a black hole. Science involves finding the best point of access available to the *explanandum*. For Hermy, that means first-person reports corroborated by more first-person reports. As Wesley Salmon wrote, “a scientific worldview is distinguished from other worldviews by the extent to which we have reason to believe it”.<sup>128</sup> And, as David Chalmers points out, we have more reason to believe qualitative facts than anything else. After all, the qualitative facts about our experiences are much more readily apparent than the biological or chemical facts.<sup>129</sup>

In fact, to say that qualitative facts can only be known from the first person point of view, and are thus distinct from chemical or biological facts, is rather odd. All knowing takes place from the first-person point of view. You can add the phrase “what it is like to...” to any experience in order to indicate the qualitative facts about that experience. There is something that it is like to know any specific fact about brain chemistry, and unless you have the experience of knowing that chemical fact yourself then there is something about chemistry that you can never know. Without a first-person point of epistemic access it is not possible to know certain facts about qualitative experience, but without a first-person point of epistemic access it is not possible to know anything at all. Surely, we should not say that facts knowable only from a first-person point of view are inherently unscientific.

McGinn might respond that qualitative facts are unscientific because they can not be *expressed* in third-person language, and therefore can not be understood except by those who are capable of placing themselves in the position of the person asserting the qualitative fact. The apparent restriction to mentalistic and demonstrative language in

expressing qualitative facts may be unique, but only because it reflects the uniqueness of the subject matter. When you are investigating qualitative properties you need to use qualitative language, just as when you are investigating astronomical properties you need to use astronomical language. You can never explain to a blind person what it is like to see red, but equally you can never explain the length of the Earth's circumference to a person unable to conceive of the Earth as a spherical object. To understand astronomical claims one must be able to take an astronomical point of view; it is par for the course that understanding qualitative claims requires one to take a qualitative point of view.

To discover (or articulate) qualitative facts Hermy might follow Nagel's suggestion, and "pursue a more objective understanding of the mental in its own right...and devise a new method—objective phenomenology not dependent on empathy or the imagination".<sup>130</sup> Nagel also says that such investigations will inevitably hit a brick wall, but he seems not to notice that his "speculative proposal" accurately describes the project of twentieth-century continental philosophy. We should look to the actual empirical work of phenomenology to assess whether there is an inevitable brick wall, not determine the explanatory limits of systematic investigation from the armchair.

Alternatively, following Patricia Churchland's suggestion, Hermy might use folk psychology as at least a beginning point for expressing qualitative facts, learning more precise means of expressing qualitative truths as she learns more about how to explain them. That is, analysis of folk psychology might provide a foundation from which Hermy can develop an objective phenomenology. One might refuse to call such investigations scientific but, to re-quote Salmon, "a scientific worldview is distinguished from other worldviews by the extent to which we have reason to believe it". It seems to me that

Hermý's process for discovering and articulating qualitative facts deserves to be called scientific if she adheres to methodological naturalism, seeks to avoid bias, establishes a system of peer review, and (to paraphrase Aristotle) provides the type and degree of precision that is appropriate to the subject matter.

Perhaps even more important than whether we should attach the label "scientific" to any systematic investigation of qualitative properties is whether qualitative facts can be explained by more traditional fields of scientific investigation. Can traditional natural sciences, such as physics, chemistry, and biology explain why the qualitative facts are what they are? After all, McGinn and Levine do not deny that we can know what the qualitative facts are (we know what they are by having the relevant qualitative experiences). They might grant that Hermý has reliable enough ways of knowing *what* the qualitative facts are, yet nevertheless deny that she can scientifically explain *why* particular brain events should be accompanied by particular qualitative states.

If Hermý can know qualitative facts by any reliable means whatsoever (even non-scientific means), and she can explain why these facts are true in light of traditional natural sciences, then we ought to conclude that there is no explanatory gap. The "gap" in "explanatory gap" refers an alleged breach between that which is known scientifically and that which is known qualitatively. If Hermý sufficiently integrates qualitative facts into the body of her scientific facts, there is no explanatory gap – even if she used divination to discover the qualitative facts in the first places.

I now turn my attention to various proposed approaches to a scientific explanation of qualitative facts. In doing so, it will be necessary to refer to examples of qualitative facts; however, I want to remain neutral about the methods and theories involved in

discovering the true qualitative claims. As it is readily accessible, I will use folk psychology as placeholder for future reliable and systematic investigation of qualitative facts. In the absence of a rigorous method of investigating qualitative properties, I will assert the same kind of qualitative facts that the folk assert. Since my goal will be to show that Jackson's epistemological intuition provides no *a priori* barrier to a scientific explanation of qualitative facts, I will be as folksy as possible to show that even if all Hermy has is a folk account of what the qualitative facts are she need not give up on her dissertation.

#### IV. Knowing *Why*: In Search of an Explanatory Model

My view is that whether Hermy can ever finish her dissertation depends on what is meant by *explanation*. As mentioned earlier, Levine argues that absent and inverted qualia thought experiments demonstrate the incompleteness of physicalist (scientific) explanation of consciousness. Torin Alter's suggestion that, upon leaving the black and white study, Mary learns facts that are not 'discursively knowable' even though they are 'about physical things' could similarly be employed against the view that qualitative facts can be scientifically explained.<sup>131</sup> After all, it is plausible that science can only explain discursively knowable facts. Daniel Stoljar's argument that Mary learns 'facts of physicalist ontology' that are not 'facts of physical theory' also suggests an explanatory gap.<sup>132</sup> In other words, virtually every recognition that the knowledge argument builds on a sound epistemological intuition is at least consistent with the existence of an explanatory gap.

The asserted barrier to scientific explanation of qualitative facts seems to be that the terms used to express qualitative facts do not connect up in the right way with the facts of physics, chemistry, and biology, such that no amount of information from the traditional physical sciences can explain why the qualitative facts are the way they are. As many see it, no amount objective third-person scientific knowledge can explain subjective first-person data. Most commonly, and somewhat distinct from the concerns raised by either Levine or McGinn, the argument that consciousness can not be scientifically explained is made on Kripkean grounds: there is no way to account for the apparent contingency of identity statements involving traditional scientific terms and terms appropriate for expressing qualitative facts.<sup>133</sup> I will, of course, need to confront such issues directly; Kripkean questions are at the center of the next chapter. In the rest of this chapter, I seek to establish a philosophical-scientific grounding for the Kripkean discussion in the next chapter.

As already mentioned, it is important to consider which model of scientific explanation Hermy uses. We might imagine that, after being assigned her topic, Hermy went to her advisors and asked them what is meant by a scientific explanation. Assuming they don't tell her that she needs to figure that out too, there are several answers that her advisors might give her, reflecting the various models of scientific explanation available in the literature. I intend to be more or less neutral on precisely which answer Hermy's advisors ought to give to the critical question of what a scientific explanation is. My claim is only that any theory of explanation that has mainstream support in the philosophy of science will indicate the direction of Hermy's research. For no adequate theory of scientific explanation are there *a priori* reasons to conclude that qualitative

facts can not be scientifically explained. Indeed, I hope to show that neuroscience has already made significant advances in the explanation of qualitative facts.

I choose to focus on five theorists of explanation as representative of five different approaches to inter-theoretic explanation: Steven Weinberg, Ernest Nagel, Patricia Churchland, Philip Kitcher, and Wesley Salmon. My goal is not only to indicate various directions Hermy might go in undertaking her research, but also to show that Jackson's epistemological intuition is inconsistent with scientific explanation of qualitative facts only in a narrow and outdated sense of 'scientific explanation'.

#### A. Final Theory Reduction

When Hermy asks her advisors what is meant by 'scientific explanation' in her assigned dissertation title, she should hope that they do not merely suggest she read Steven Weinberg's *Dreams of a Final Theory*. First of all, it would devastate Hermy's hopes of ever graduating. According to Weinberg, a reduction of all scientific knowledge to a comprehensive and unifying theory of physics is the ideal of a complete scientific explanation. 'Reduction' means many different things to different people and it is important to keep the different senses of it clear. After all, Hermy's work is inherently inter-theoretic, in that she appeals to traditional levels of scientific analysis (physics, chemistry, biology) in order to explain why the facts at another level of analysis (the qualitative level) are the way they are. Weinberg suggests a model of explanation in which each scientific level of inquiry is explained by showing that the properties referred to are really nothing more than the properties known at lower levels of investigation and are ultimately nothing more than the properties of physics. In the end, there is "only



physics". According to Weinberg, the "final theory" is a theory of physics that makes every theory in every other science redundant and thereby obsolete.

The feature of this model of explanation that would ensure that Hermy never finishes her dissertation is ontological simplification. According to Weinberg's model of explanation, sciences other than physics have only a heuristic value in helping ultimately to discover the final theory. On such a view, there simply can not be such things as qualitative states that can only be known by means of one's own qualitative states. To claim that qualitative facts ultimately disappear in a whirlwind of explanatory physics is contrary to Jackson's epistemological intuition, which asserts that their qualitative properties can be known only by means of qualitative experience (and thus can not be known by means of physics). If qualitative facts are to be explained, Hermy will need a model of scientific explanation that does not require ontological simplification. Qualitative states will have to be taken seriously as a phenomenon in their own right.

The second reason that Hermy should hope she is not directed towards a Weinberg-style of explanation is that Weinberg has few fans in the philosophy of science community.<sup>134</sup> Few (if any) historical examples of reduction meet his stringent standards of explanation. The terms used in theories of chemistry and biology refer to real phenomena. Though particle physics no doubt studies the natural world at a more fundamental level than biology, it is not the case that biology has no proper domain of its own. Physics could never make biology obsolete; it is no threat to point out (as Jackson's epistemological intuition does) that physics can never make phenomenal psychology obsolete either. Most importantly, if Hermy is told by her advisors to pursue a Weinberg-style explanation then she is being given advice that does not reflect the current

understanding of scientific explanation among philosophers of science. Many philosophers (influenced by Kim and Chalmers) do hold a metaphysical view that assumes micro-reductive view of scientific explanation; however, such views are not held for reasons pertaining to the nature of scientific explanation per se. Such views are discussed in Chapters Two and Four. The question here is epistemological, and Churchland, Kitcher, and Salmon all, in various ways, argue that actual scientific practice is not micro-reductive. For now, notice that in order to conclude *a priori* that Hermy could not scientifically explain her vast qualitative knowledge one would have to consider ontological simplification a necessary part of a scientific explanation.

#### B. Nagel-Reduction

We should also question whether Hermy is working with a current understanding of scientific explanation if she is directed primarily to Ernest Nagel's classic account of a reductive explanation. According to Nagel, a reduction is a type of intertheoretic explanation that satisfies Hempel's deductive-nomological (D-N) account of explanation in which the *explanandum* is itself a scientific law.<sup>135</sup> Working on the basis of such a model, Hermy would have explained the laws regulating qualitative facts if those laws are the conclusions of deductive arguments with premises that include at least one general law at a lower level of investigation (say, biology) and any necessary bridge laws. Bridge laws are necessary to connect the terms of the reduced theory to the terms of the reducing theory, in order to make deductive arguments connecting the two theories possible. Bridge laws connect the terms used in the reduced theory to the terms of the reducing theory. For instance, 'heat is mean molecular motion' defines a term of

thermodynamics in the terms of classical dynamics. Since reductive explanations are assumed to be deductive arguments, no conclusion about thermodynamics can follow from premises established by classical dynamics unless there is an additional premise expressing the relationship between the terminologies of the two sciences.

Nagel's requirements are a barrier to Hermy finishing her dissertation. Most fundamentally, it is not clear that there are law-like propositional expressions of qualitative facts that can serve as the conclusion to a deductive argument. Quite plausibly, qualitative facts are too nearly ineffable to realistically be deductively implied. I also doubt that Hermy would ever establish any bridge laws between qualitative terms and neurobiological terms. I am tempted to concede that Hermy could never provide a Nagel-reduction of qualitative facts, point out the abundance of historical and hypothetical counter-examples to D-N explanation generally, as well as Nagel-reductions specifically, and move on. However, it is worth noting that Nagel's position is more subtle than commonly recognized.

Nagel distinguishes between two kinds of bridge laws that are involved in reductive explanations. In the first, "bridge laws specify conditions for the occurrence of an attribute".<sup>136</sup> These conditions may be necessary for the occurrence of the attribute, sufficient for the occurrence of the attribute, or both. For example, "a familiar bridge law states that a gas has a certain temperature when the mean kinetic energy of its molecules has a certain magnitude".<sup>137</sup> In other words, having a certain temperature is one thing; the conditions under which we say that a gas has that certain temperature is another. There are still two things involved in that, "the class of things connoted by a predicate in a reduced law may indeed be quite different from the attribute connoted by the predicates

of the reducing theory”.<sup>138</sup> The second kind of bridge laws, “consist in showing that things and processes initially assumed to be distinct are in fact the same”.<sup>139</sup> A familiar example is of the reduction of physical optics to electromagnetic radiation. In this case what was previously presumed to be two distinct things (light and electromagnetic radiation) are shown to be one.

The difference between the two kinds of bridge laws is ontological simplification. After accepting the second kind of bridge law, we have simplified our ontology, but after accepting the first we have not. The critical point is that there can be *explanatory reduction* (which is provided by either kind of bridge law) without *ontological simplification* (which is provided in only the second case). Whether a bridge law serves the appropriate role in an explanation of the reduced theory is an empirical question. But once the empirical work is done there is the further, and non-empirical, question whether the bridge law justifies a simplification of ontology. An example of the distinction between the two kinds of bridge laws can be found by examining the demise of the nineteenth-century demonic possession theory of epilepsy. When empirical science showed that epileptic seizures are associated with abnormal cortical-electrical activity there was a metaphysical choice to be made. Nothing in the empirical evidence itself demanded ontological simplification; neurologists might have concluded that abnormal cortical-electrical activity specified “conditions for the occurrence” of demonic possession. They chose (on humane, not empirical grounds) to eliminate the theory of demonic possession and conclude that ‘seizure’ and ‘abnormal electrical activity’ refer to one and the same thing.<sup>140</sup>

Given that the micro-reductionist's dream of ontological simplification is inconsistent with Jackson's epistemological intuition, the recognition that Nagel-reductions can be achieved without ontological simplification gives Hermy some reason to be hopeful. Having discovered the best way to formulate the qualitative facts, Hermy need not define (or translate, as it is often put) the terms used to express qualitative facts in the language of the traditional natural sciences. It would be sufficient were she to specify, in the terminology of the traditional natural sciences, necessary and/or sufficient conditions for the occurrence of, say, a red qualitative experience. There may be insurmountable barriers to Hermy's attempts to achieve a Nagel-reduction of qualitative facts (especially if certain views about multiple realizability turn out to be true), but those are empirical questions to be sorted out in the laboratory – not the armchair.<sup>141</sup>

### C. New-Wave Reduction

If Hermy is advised to read Patricia Churchland, and seek a new-wave reductive explanation of qualitative facts, she will certainly have been given a more up-to-date model of scientific explanation. Yet, conventional wisdom suggests that even then she will surely never graduate. All too often, Churchland's eliminativism is taken to mean that qualitative facts are eliminated in favor of a purely neurobiological explanation, and Hermy has been told that she can not deny the existence of qualitative facts. However, 'elimination' is a term that applies to theories and theoretical concepts, not to phenomena themselves. Specifically, Churchland's eliminativism is an empirical prediction that the theory of folk psychology will be eliminated (it's terms will be replaced) by a future theory of scientific psychology that will stand in a reductive explanatory relationship with

the traditional physical sciences. 'Eliminativism' does not mean that qualitative states are illusory, but only that our folk theories about qualitative states are importantly mistaken.<sup>142</sup> Churchland responds to the objection that neurobiology can't explain everything about consciousness not by highlighting our limited capacity to imagine the explanatory power of future science. Instead she says that the objection is "straw through and through".<sup>143</sup> She even concedes "with the advantage of hindsight" that the term 'eliminative materialist' is "an invitation to misunderstanding" in that it easily leads to the assumption that she is advocating an elimination of higher-levels of investigation.<sup>144</sup>

Patricia Churchland modifies the D-N model of explanation. Churchland's most significant contribution is to acknowledge that most often in order for the deductive argument to be formed both the reducing theory and the reduced theory need to be modified and improved. The mechanism of this modification is the "co-evolution of theories". That is, reductive contact between a lower-level theory and a higher level-theory involves back and forth cooperation, in which insights and discoveries at the higher and lower levels of investigation serve to "correct and inform one another". Highlighting the role of top-down (as well as bottom-up) constraints she writes that, "psychology and neuroscience should each be vulnerable to disconfirmation and revision at any level by the discoveries of the other" and that "neuroscience and psychology need each other".<sup>145</sup> Responding to the objection that microscopic investigations can only explain a small part of human life, and inevitably leave out "the rich matrix of relations (the human creature) bears to the other humans, practices and institutions of its embedding character", Churchland and Churchland write that, "the proper response to this objection is to embrace it".<sup>146</sup> Any reductive account of the mind-brain must explain

how the brain is able to take into account its complex cultural milieu. Recognizing that culture is a “major challenge” for neurobiology, she chastises neuroscientists for not recognizing the importance of accounting for love, moral character, and hidden motivations.

Patricia Churchland likens the co-evolution of theories to “two rock climbers making their way up a wide chimney by bracing their feet against the wall, each braced against the back of the other”.<sup>147</sup> Elsewhere, she analogizes the process of co-evolution to, “a long and slowly maturing marriage”.<sup>148</sup> On Patricia Churchland’s view, neuroscience and psychology can work together to achieve a new-wave reductionist explanation of consciousness only if each science is constrained by the other (holds to no theory that is disconfirmed by the other science) and is inspired by the other (looks to the other science for refinement of *explanandum*, problem sets, and argument patterns). I quote Patricia Churchland more extensively than would normally be necessary because the difference between the caricatures of eliminative materialism that one often finds and the position Patricia Churchland actually defends is astonishing. To accuse Patricia Churchland of not taking consciousness seriously is simply a misreading.

In fact, Hermy could look to Churchland for guidance both in the discovery of qualitative facts and in the scientific explanation of qualitative facts. Though Churchland predicts that eventually folk psychological theories will be replaced by a scientific psychology, she also argues that folk psychology has a critical role to play in getting scientific psychology off the ground. In essence, folk psychological theories provide the initial set of conceptual apparatus and testable predictions needed to ground the research from which a scientific psychology can emerge. Folk psychology can provide Hermy

with a starting point for clearly identifying her *explanandum*. Hermy's emerging scientific psychology can then increasingly constrain and inspire her emerging neuroscience as the two pursuits move towards "reductive consummation".<sup>149</sup> And even though it is now twenty years old, Churchland's *Neurophilosophy* persuasively makes the case that, far from being impossible, the foundation for Hermy's dissertation is already being laid.

#### D. Unificationist Explanation

If, upon asking what is meant by 'scientific explanation' in her assigned dissertation title, Hermy is advised to read Philip Kitcher and to seek a unificationist explanation of qualitative facts, then her advisors might have already concluded that there is no reduction of qualitative facts to the facts of traditional physical sciences. Not all models of intertheoretic explanation are versions of reduction.

According to Philip Kitcher, if we want to explain some given phenomena we need to show that it follows from the general argument forms which are contained in the "set of arguments which best unify" the collection of scientific beliefs.<sup>150</sup> This model is reductionist in that it seeks a reduction to the fewest number of general argument forms possible, showing that particular events are to be expected on the basis of general, and more basic, laws. However, it is not reduction in the sense of higher-level theories following deductively from lower-level theories (with the aid of bridge laws, if necessary). For Kitcher, conceptual and logical unity is sufficient for inter-theoretic explanation. A higher-level theory can be explanatorily unified with a lower-level theory without the higher-level theory being logically entailed by the lower-level theory. Though



Kitcher no longer accepts unification as an overall model of explanation, unification is still widely recognized as a worthy goal within a larger, multi-faceted view of explanation.<sup>151</sup>

For an example of explanatory unity without Nagelian reduction, Kitcher points out there has been no reduction of classical genetics to molecular genetics, and none seems to be forthcoming.<sup>152</sup> Even if we were to establish bridge laws between the two fields, linking the vocabulary of classical genetics to molecular genetics (itself an impossible task), “simply plugging a molecular account into the narratives offered at the previous stages would *decrease* the explanatory power of those narratives”, because it would be too complex to provide the intellectual satisfaction associated with the classical account.<sup>153</sup>

Yet, molecular genetics “has done something important” for classical genetics.<sup>154</sup> Kitcher mentions three such things, pointing to three explanatory achievements of inter-theoretic research that are independent of classic reduction. 1) By describing the mechanism of replication, molecular genetics has provided an *account of something that is presumed* (and apparently unexplainable) at the level of classical genetics. 2) With accounts of things like mutant alleles, molecular genetics has provided a *conceptual refinement* of the classical genetic notion of mutation. 3) Most importantly, because molecular genetics can connect phenotype to genotype, molecular genetics provides an *explanatory extension* of classical genetics in that the connection of phenotype to genotype demonstrates that, “there is some problem-solving pattern of [classical genetics] one of whose schematic premises can be generated as the conclusion of a problem-solving pattern of [molecular genetics]”.<sup>155</sup> However, it does not follow that the

derivation from molecular genetics has more explanatory power than the derivation from the classical theory, again due to the complexity of the molecular account. In short, understanding phenotypic manifestation of a gene requires “constant shifting back and forth across levels”.<sup>156</sup>

Hermey might take the explanatory relationship between classical and molecular genetics as her guide in exploring the role of neurobiology for our understanding of qualitative facts. It might be objected that Hermey’s finished dissertation would then, at best, provide a partial explanation of qualitative facts. Hermey would fail to show that the facts of the traditional natural sciences make the qualitative facts true. But that is just the point of Kitcher’s analysis of the relationship between classical and molecular genetics. Classical genetics can not be Nagel reduced to molecular genetics, but classical genetics can be brought into explanatory unification with molecular genetics. If Hermey can place the qualitative facts within the explanatory nexus (Kitcher’s term) of her beliefs derived from the traditional natural sciences, then it is only fair to say that there is no explanatory gap, the scientific worldview does not leave out the qualitative aspect of experience.

It is not hard to find examples where the relationship between folk psychological accounts of qualitative facts and neurobiology stand in a relationship parallel to Kitcher’s account of the relationship between classical and molecular genetics. Regarding neuroscience’s capacity to provide an account of something that is presumed in folk psychology: stress causes fatigue. From time immemorial, Grandma has known that prolonged subjective experiences of stress causes subjective experiences of fatigue (though Grandma tended to leave out ‘subjective experiences of’). To folk psychology the idea that stress causes fatigue is so obvious that it needs no explanation, and folk

psychology is certainly not prepared to describe the mechanisms by which stress causes fatigue. Neurobiology, however, provides a hormonal account of folk psychology's presumed connection between stress and fatigue. Second, regarding neuroscience's ability to provide a *conceptual refinement* of qualitative concepts, the folk psychological concept of "attention" is exceedingly vague. Neuropsychology has shown that the folk psychological concept of attention actually refers to a variety of related phenomenon.<sup>157</sup> Of course, the splintering of the folk psychological concept of attention is one of Patricia Churchland's standard examples of a folk psychological concept that is being splintered and is thus up for elimination.<sup>158</sup>

Third, and most important, is the point to be made regarding neuroscience's ability to provide an *explanatory extension* of folk understanding of qualitative states. Since folk theories are not expressed with the kind of precise propositions suitable for deductive argumentation, my reliance on folk psychology is most restrictive here. Nevertheless, it is helpful to discuss neurobiological correlates of folk observations regarding red. In some cultures, colors have different emotional significances; white is the color of death in Asia but of purity in the West. Giving red flowers instead of yellow ones can have an emotional (and qualitative) impact on either the giver or receiver of flowers. Hermy will know the mechanism by which colors take on particular emotional significance. Plausible speculation suggests the emotional significance of color correlates with activation of neural connections between V4 and the limbic structures, along with the linguistic mechanisms associated with processing cultural information. If Hermy sorts out the details of these connections she will be utilizing a schematic argument (in neuroscience) that establishes a premise which is useful in explaining qualitative facts.

## E. Causal-Mechanical Explanation

Hermi should be most optimistic if her advisors' direct her primarily to Wesley Salmon's model of scientific explanation, particularly if they highlight Salmon's advice to archaeologists seeking scientific credibility for their discipline.<sup>159</sup> Salmon observes that many archaeologists equate a scientifically credible explanation with the achievement of a D-N explanation, and that unjustified standards of scientific explanation have become barriers to scientific progress in the field. Salmon points out that the D-N model hasn't been the official view in philosophy of science for a long time and that other models of explanation (especially his own, of course) are likely better suited to archaeology. It seems to me that Salmon's advice to archaeology equally applies to qualitative psychology.

In contrast with all of the models so far discussed, Salmon contends that scientific explanations are not arguments. Instead, explanation involves delineation of causes. If you want to know why an event happened, it is sufficient to determine the cause of the event. As Salmon likes to put it, he wants to put the 'cause' back into 'because'.<sup>160</sup> Strong intuitions favor a causal account of explanation. After all, it is natural to answer the question "why did event *x* happen?" with "it was caused by *y*". Of course, as David Hume argued, causal connections are not empirically accessible: observation can be made only of constant conjunctions, not of necessary connections. Many interpret Hume to mean that causation is an illusion, something provided by the imagination.<sup>161</sup> Hume's problem (as Salmon sometimes calls it) is to distinguish empirically between mere correlations and genuine causal relationships. Given the intuitive strength of the causal

approach to explanation, the model probably ought to be accepted or rejected depending on whether Salmon sufficiently addresses Hume's problem.

Salmon does not balk; he claims to solve Hume's problem. He points out that Hume sought a logical or metaphysical connection between separate events. In contrast, Salmon understands causation as a physical process which transmits marks throughout the process. Because of these two points of departure from Hume (processes instead of events, physical connection instead of logical or metaphysical connection), Salmon can claim that, "on the view of causality I am advocating, causal connections exist in the physical world and can be discovered by empirical investigation".<sup>162</sup> One particular example demonstrates not only why Salmon considers causation to be an empirically discoverable physical process, but also how he distinguishes genuine causal processes from pseudo-processes. "Consider a rotating spotlight, mounted in the center of a circular room, which casts a spot of light on the wall. A light ray traveling *from* the spotlight *to* the wall is a causal process; the spot of light moving around the walls constitutes a pseudo-process".<sup>163</sup> The light traveling from the spotlight to the wall is causal because "a genuine *causal process* is one that can transmit a mark; if the process is modified at one stage the modification persists beyond that point *without any additional intervention*".<sup>164</sup> For instance, imagine that the light is white initially; the light on the wall will be white. If the light is fit with a red filter, the information in the light will be changed (it will have a different mark) and the light on the wall will also be red.<sup>165</sup> In contrast, the light passing across a particular imaginary line on the wall is a pseudo-process. If you were to put a red spot on the otherwise white wall then it would affect the appearance of the light only as long as the light was still focused on the red-colored part of the wall. In the former case,

the mark is transmitted; in the latter it is not. Therefore, the former is a genuine causal process and the latter is not. Other examples of pseudo-processes include the motion of a horse on a movie screen (as opposed to the genuinely causal processes of either the motion of the actual horse who was filmed for the scene, or the light sent from the projector to the screen).

Salmon's position addresses Hume's concern that only constant conjunction, and not the necessary connections asserted by causal claims, can be discovered empirically. After all, if we interpret the cases in Humean terms, there is constant conjunction between both 1) the light in the spotlight and the light on the wall and between 2) the series of light spots moving around the circular room. Yet, on the basis of observable differences it is possible to distinguish which of these constant conjunctions involves a genuine causal relationship and which involves a mere correlation. Notice that Salmon's theory of mark transmission does not take the spotlight as one event, and the appearance of the light on the wall as a second event. They are part and parcel of a single causal process. Thus, instead of saying Salmon solves Hume's problem, it might be more accurate to suggest that he keeps it from arising in the first place. When the unitary process is not divided into distinct events, the question of the connection between the events simply does not arise. The only task remaining is distinguishing real from pseudo process – a task accomplished with the notion of mark transmission. One might respond to Salmon's criteria for distinguishing between genuine and pseudo causal processes by arguing that transmitting a mark seems like a sufficient condition for a genuine causal process, but unless all causal processes transmit marks we cannot say that Salmon has solved Hume's problem. Perhaps Salmon has not solved Hume's problem entirely, but if

mark transmission is a sufficient condition for genuine causal processes, Salmon has at least demonstrated that some causal claims (perhaps ones involving phenomenal psychology) are scientifically respectable.

We already know a great deal about what Hermy's dissertation will say if she relies on Salmon's model of explanation in her dissertation. When we ask "why is it like *that* when we see something red?", the answer relates a process by which light of a certain wavelength hits the retina, is picked up by certain cones, is translated into chemical and electrical signals which are sent via the optic nerve to the occipital lobe (V1 specifically), and from there move across the cortex in an anterior fashion through V2, V3, V4, and so on. The end of the process is the qualitative experience of what it is like to see red. Of course, this account isn't going to satisfy any property dualists—after all, they are interested in why these processes should produce *that* particular what it is likeness. But, insight into why particular processes produce particular qualitative effects should not be asked of generalized sketches of possible future science. When the details are in we can argue about how much they illuminate the nature of conscious visual experience itself. I'm betting that once we understand how certain processes produce qualitative states *at all* we will thereby largely understand how such processes produce the *particular* qualitative states that they do.

The only question remaining is whether these correlates of visual experience involve causal processes or pseudo-processes. Clearly, chromatic vision is not a pseudo-process. Just as with Salmon's example of the red filter fitted to the spotlight, marks are transmitted at each stage of the process. Clinical neuropsychologists are well aware of the impact of various modifications along each stage of the process. If the relevant cones are

missing, the visual experience will be achromatic. If bilateral damage to V1 exists while the rest of the system is intact, the phenomenon known as “blindsight” might emerge. If one cortical hemisphere has been removed (to treat seizures, for instance) bilateral hemianopia will result. And so on. The critical point for Hermy’s dissertation is that *the mark is transmitted all the way to the conscious visual experience itself. What it is like to see red is the end result of the causal process, and thus, on Salmon’s model, explained by the causal process.* Since the above sketched causal process is already scientifically well established, Hermy’s dissertation might be finished much sooner than her advisors would have guessed.

Explaining conscious experience by pointing to neurobiological correlates is generally dismissed as being a demonstration of a mere correlation. But if we accept Salmon’s grounds for distinguishing between mere correlations (pseudo-processes) and genuinely causal processes, there is no justification for this critique. Salmon’s account of mark transmission justifies the claim that the neurobiological account provides a *causal explanation* of consciousness. The mark in the final state being the very lived character of the experience, the qualitative property of the experience. The qualitative experience includes the transmitted mark in that the changes observable in earlier stages of the process alter the subjective character of the visual experience. Given that Hermy works out all the neurobiological correlates involved in every aspect of seeing red (the perceptual, the emotive, the cultural etc.), there is no *a priori* barrier to her finishing the dissertation. After all, there is no reason to think there will be a failure of mark transmission regarding any of these aspects. There is the separate issue regarding how she describes the phenomenal experience itself, but marks transmitted from the



neurobiological level are even expressed in folk understandings of qualitative properties. Hermy's dissertation will be long and complicated, but claims that there is something left out (the redness of seeing red, as Chalmers puts it) lose their argumentative force.

## V. The Knowledge Argument and the Explanatory Gap

In "Causality and Explanation", Wesley Salmon argued that unificationist and causal models of explanation can each sometimes offer an explanation of a given phenomenon. In such cases, the two explanations can be different and yet compatible; the causal and unificationist explanations are not the same, but neither are they mutually excluding. Near the end of that essay, he makes a brief reference to his former graduate student Gemes, who suggested that "it is futile to try to explicate the concept of scientific explanation in a comprehensive manner. It might be better to list various explanatory virtues that scientific theories might possess, and to evaluate scientific theories in terms of them".<sup>166</sup> Though Salmon mentions only unification and delineation of causes among explanatory virtues, it seems more than reasonable to recognize that new-wave reduction is also an explanatory virtue.

In other words, the best approach to explaining consciousness is not to first choose an explanatory model and then do research that fits only the chosen explanatory model. On the contrary, research should proceed recognizing by the value of unification, causal delineation, and reduction. Insights and discoveries made along the way contribute to an explanation of consciousness to the extent that they fit any viable model of scientific explanation. Naturally, this contribution only enhances the prospects of a scientific explanation of consciousness, as it seems likely that certain aspects of

consciousness will be explained reductively, other explained causally, and still others explained through unification with other scientific knowledge.

Nevertheless, scientific explanation of qualitative subjective experience may turn out to be impossible. The brain may turn out to be too complicated for even Hermy's extraordinary human mind (as McGinn suggests). Consciousness may turn out to be some as yet undiscovered fundamental property of the universe (as Chalmers suggests). We can not even rule out the possibility that there really is some immaterial soul stuff (as Descartes suggests). Also, assuming that Hermy does finish her dissertation, I am not suggesting that any of us normal people would have either the time to read it (we have only a few decades to read) or that it would provide us the kind of intellectual satisfaction commonly associated with scientific explanation (we are not nearly as smart as Hermy and might have the same reaction to Hermy's dissertation that my six-year-old has when reading mine). To grasp all the details simultaneously and have that "aha!" feeling might be quite beyond our limited capacities. Nevertheless, *a priori* speculation concerning what can be determined on the basis of some futuristic or idealized body of scientific knowledge can not justify pessimism. Specifically, there is nothing about the narrow reading of Jackson's epistemological intuition that, in itself, makes Hermy's project impossible. On the contrary, though a scientific explanation of qualitative facts may turn out to be impossible, such a claim would have to be made empirically. Armchair speculations about Mary and Hermy do not demonstrate the existence of an explanatory gap.

It is unfortunate that the debate in the philosophy of mind literature about the explanatory gap focuses almost exclusively on overly simplistic and outdated

understandings of the nature of scientific explanation. The lesson to be drawn from contemplating Hermy is that it matters what a “scientific explanation” is taken to be. Naturally, we hope that Hermy’s dissertation research will be based on the best models of scientific explanation that are available, the models that are currently viable in mainstream philosophy of science literature, and the models that are best suited to Hermy’s subject matter. Unfortunately, there is very little discussion of the subtleties of new-wave reduction, unificationist explanation, or causal explanation in the literature on qualitative experience.<sup>167</sup> And the discussion of reduction almost invariably passes over any subtlety offered by new-wave approaches to reduction. Discussion of the explanation of consciousness tends to operate within the context of Nagel-reduction or micro-reduction. Yet, unificationist and causal models of explanation were both developed in recognition of compelling historical and hypothetical counter-examples to D-N explanation and emerged as philosophy of science became more interested in biological explanation. Given that D-N models, by consensus, do not capture the practice of actual scientists, and that the brain is a biological organ, it is natural for Hermy’s advisors to appeal to precisely those models that were developed from a historical understanding of the biological sciences.

## Conclusion

Ontologically simplifying models of explanation have dominated the philosophy of mind literature because Kripkean accounts of *a posteriori* necessity are ontologically simplifying. Jackson’s response to the central arguments of my last two chapters would be that I failed to notice the *a priori* story there is to tell about *a posteriori* truths. On

Jackson's view, facts that are not explicitly physical can only have a place in physicalist ontology if they are *a priori* entailed by explicitly physical facts. That is, Jackson would respond that I have been mis-interpreting premise four of the knowledge argument ("if physicalism is true then there are no facts other than all the physical facts") and that physicalism is committed to a much stronger sense of the completeness of the physical facts than I have recognized.

---

Notes to Chapter Three

118 See McGinn, *The Mysterious Flame* and McGinn, "Solve Mind Body Problem."

119 Levine, "On Leaving Out What It's Like," 543.

120 Of course, the modal part of Jackson's argument is not limited to the first premise and can only be understood in light of premise four. In chapter Two I suggested the following formulation of Jackson's modal intuition: "it is possible to know all the facts about a minimal physical duplicate world without chromatic experience". The anti-physicalist conclusion of the knowledge argument requires both the modal and epistemological intuitions.

121 Notice that this formulation refers to 'qualitative' experience instead of the narrower 'chromatic experience'. Uncontroversially, Jackson is using chromatic experience as an example of qualitative experience generally. Mary might just as well have been locked in a sound-proof room and forced to learn all the physical facts about auditory experience.

122 Nagel, "What is it Like to be a Bat?," 325.

123 Jackson asks the reader to make similar assumptions for similar reasons in Jackson, "The Case for *A Priori* Physicalism" and Jackson, "A Priori Physicalism".

124 Chalmers, *The Conscious Mind*, xii.

125 McGinn, *Mysterious Flame*, xi.

126 Levine, "Leaving Out What It's Like," 546.

127 Popper, "Conjectures and Refutations," 4.

128 Salmon, "Causation and Explanation," 77.

129 Chalmers, *The Conscious Mind*, 3. Of course, the idea that consciousness is more readily apparent to us than anything else dates back at least to Descartes.

130 T. Nagel, "What is it Like to be a Bat?," 329.

131 Alter, "Limited Defense of Knowledge Argument," 50.

132 Stoljar, "Two Conceptions of Physical," 312.

133 See Kripke, *Necessity and Identity*, 151.

134 See Salmon, "Dreams of a Famous Physicist," for an extended discussion of the ways in which Weinberg has overlooked or misunderstood much of the history of the philosophy of science.

135 The D-N model of explanation is most closely associated with the work of Carl Hempel. According to Hempel, a fact has been explained when the proposition stating the fact is the conclusion of a deductive argument that includes among its premises at least one general law and at least one specific observation. For instance, the general law that metal rods expand when heated along with the specific observation that

---

this rod has been heated explain why this rod has expanded. See Hempel, "Two Basic Types of Explanation."

136 Ernest Nagel, "Issues in Reduction," 914.

137 Ibid, 914.

138 Ibid, 914.

139 Ibid, 914.

140 The horror of the seizure remains; there is still the striking observation of someone's mind going blank and their body being taken over by some dark and internal force. The empirical evidence supporting the demonic possession theory did not go away. But eliminativism, in this particular case, refocused human understanding of people suffering from epilepsy. No longer did they have the devil inside them.

141 For an excellent account of how of the multiple realizability problem and strategies for responding to it, see Polger, *Natural Minds*, 5-29.

142 According to Churchland, much of folk psychology has already been proven wrong in that the concepts of folk psychology are being splintered: what folk psychology considers a single thing has been proven by neuroscience to be a range of distinct phenomenon. The folk psychological concepts of awareness, memory, and pain are her most frequently cited examples.

143 Churchland and Churchland, "Intertheoretic Reduction," 253.

144 Patricia Churchland, "Do We Propose to Eliminate Consciousness?," 298. There is a significant difference between Patricia's work and Paul's earlier work. For example, in his 1981 "Eliminative Materialism and the Propositional Attitudes" Paul Churchland imagines a future in which not only is folk psychology eliminated but so is all intentional communication (in favor of the type of communication that happens between one's right and left hemisphere).

145 Churchland, *Neurophilosophy*, 376, 373.

146 Churchland and Churchland, "Intertheoretic Reduction," 251.

147 Churchland, *Neurophilosophy*, 373.

148 Churchland and Churchland, "Intertheoretic Reduction," 253.

149 Churchland, *Neurophilosophy*, 283.

150 Kitcher, "Explanatory Unification and Causal Structure," 434.

151 For a critique of Kitcher see Jones, "Reductionism and the Unification Theory of Explanation," 24. For an example of how unification continues to nevertheless play a role in explanation see, Hardcastle, "Reduction, Explanatory Extension," 419.

152 Kitcher, "1953 and All That," 335.

153 Ibid, 348.

154 Ibid, 368.

155 Ibid, 364.

156 Ibid, 371.

157 Kolb and Wishaw, *Fundamentals of Neuropsychology*, 578-603.

158 Patricia Churchland, *Neurophilosophy*.

159 Salmon, "At-At," 193.

160 See Salmon, "The Importance of Scientific Understanding", 89 and Salmon, "A Third Dogma of Empiricism," 107.

161 I have always thought this is too strong of a reading of Hume. While Hume certainly points out that the causal connection between two events can not, itself, be observed, it goes too far to suggest that causation is an illusion and not a property of the world. On my reading, Hume is suggesting that the belief that a causal connection exists in any particular case must be supplied by the faculty of reason. Causation is a relation that does exist "out there", describing what actually pertains between billiard balls. But our

---

knowledge of the relationship is inferred from observation of constant conjunctions, plus something else that is difficult to precisely determine. See Hume, *A Treatise of Human Nature*.

162 Salmon, "A New Look at Causality," 23.

163 Salmon, "At-At," 194.

164 Salmon, "Four Decades of Scientific Explanation," 108.

165 Ibid, 108.

166 Salmon "Causation *and* Explanation," 78.

167 See Hardcastle, "Reduction, Explanatory Extension," for an exception.

## Chapter Four:

### Entailment and the Placement of Qualitative Facts

When Jackson changed his position in 1998 he only abandoned the claim that there are facts other than physical facts (premise three of the knowledge argument). Even after 1998 Jackson still maintains that, if physicalism is true, there are no facts other than physical facts. Put differently, Jackson has consistently held that philosophers must choose either reductive physicalism or some version of dualism. A consistent theme of this dissertation is that Jackson offers a false dichotomy when he insists that these are the only two options. Jackson's response to both Chapter Two and Chapter Three of this dissertation would be that I have understated the sense in which physicalism is committed to completeness. In answering this anticipated objection, I will be critiquing Jackson's current physicalist position, as well as his former anti-physicalist position. Though Jackson was right to abandon the knowledge argument, he abandoned the wrong part.

#### I. The Entailment Thesis

The central concern of Jackson's recent writings is to argue that a metaphysical system is only complete if all the facts about our world that are not explicitly about the fundamental entities, properties, and relations countenanced by the metaphysical system are *a priori* entailed by the fundamental facts. Substance dualists must hold that all facts are included among, or entailed by, the facts about fundamental mental and physical facts. Idealists must hold that all facts about our world are among, or are entailed by, the

fundamental facts about ideas. Similarly, physicalists must hold that all of the facts about our world are either among, or are entailed by, the fundamental physical facts. As explained in the second part of the chapter, whether I agree with Jackson depends on what is meant by ‘entailment’.

#### A. Linguistic Entailment and the Placement Problem

According to Jackson, any serious metaphysic faces the *placement problem*; that is, every serious metaphysical system faces the problem of finding a place within the metaphysical system for those properties that are not explicitly among the privileged ingredients of the system. Therefore, physicalists must find a place for, among other properties, the psychological properties of the world. After all, physicalism (like any serious metaphysical position) claims to provide a complete account of the world in terms of a limited number of ingredients. Jackson persuasively argues that physicalists are thereby committed to the following supervenience thesis: “any world which is a *minimal* physical duplicate of our world is a duplicate *simpliciter* of our world” (I refer to this as the *minimal supervenience thesis*). As explained in both the Introduction and Chapter Two, a minimal physical duplicate world is what you would get if you created a duplicate of all the instances of physical properties and relations of our world and added nothing else. Of course, this supervenience claim does not solve the placement problem. It just reasserts that the placement problem must have some solution.

In a further claim, Jackson holds that if the minimal supervenience thesis is true, then there must be an *a priori* entailment of any given psychological fact from the complete set of physical facts. In other words, Jackson believes that the only solution to



the placement problem is *entry by entailment* (the *entailment thesis*): the facts about all the properties of the world must be entailed *a priori* by the facts about the privileged list of ingredients in the metaphysical system. His primary example is that, if physicalism is true, any given psychological fact (indeed, any fact at all) must follow deductively from the complete set of physical facts. If there are any facts about the world that are not deductively entailed by the physical facts, then physicalism is false. As Jackson puts it, there must be an *a priori* story to be told (in principle) about how the physical facts entail the psychological facts. It is important to note that the “stories” Jackson refers to are not the kind of stories that humans can actually tell; they are purely idealized linguistic constructions. The actual practice of science can only approximate them because actual science is presented in actual, not idealized, language.

In 2001 Jackson and Chalmers co-authored a clarification of the entailment thesis. Jackson and Chalmers explicitly avoid the question of whether the microphysical facts *actually* entail the psychological facts (a question about which they now disagree). They restrict themselves to the assertion that, if physicalism is true then such entailments must exist in principle (an assertion about which they agree). Of course, Jackson and Chalmers are not suggesting that such entailments can be expressed in practice. Instead, they are committed to a linguistic entailment in which sentences that can in principle be constructed about one set of properties entail the sentences that can in principle be constructed about another set of properties. As they explain their position, the conjunction of *P* and *T* must entail *M*. *P* stands for “the conjunction of microphysical truths about the world,” including all of the true propositions about the fundamental entities and properties of a completed physics.<sup>168</sup> *P* is also stipulated to include the

conjunction of all the laws of completed physics. In other words, *P* is an extremely long sentence that could never be constructed or understood in practice. It is the story about microphysics that can only be told by an idealized observer. *P* might be summarized as God's understanding of the microphysical world. Paralleling Jackson's use of 'minimal', *T* is stipulated by Jackson and Chalmers to be a "that's all" clause. So, according to Jackson and Chalmers, the conjunction of *P* and *T* excludes statements about non-physical entities such as angels or ectoplasm.<sup>169</sup> Jackson and Chalmers further stipulate *M* to be any give single statement that expresses "a typical macroscopic truth concerning natural phenomena".<sup>170</sup>

They deliberately exclude psychological phenomena from the domain of *M* in order to avoid controversies about whether physicalism is true (a claim about which they now disagree, starting in 1998) and focus on what physicalism entails (about which they have consistently agreed). In order to stick to issues about which they agree, instead of psychological facts, they focus on claims about such things as water and life.

Nevertheless, as physicalists who eschew eliminativism must place qualitative facts within their ontology just as much they must place biological facts, when Jackson and Chalmers assert that physicalism is only true if the conjunction of *P* and *T* entails *M*, they have asserted a general claim about the placement problem for physicalism. Any fact about the world that is not included in *PT* must be entailed by *PT* if physicalism is to be a complete metaphysical system. Recall that, on my formulation, the fourth premise of the knowledge argument asserts that "if physicalism is true there are no facts other than the physical facts". The entailment thesis clarifies this fourth premise: according to Jackson, if physicalism is true then there are no facts other than those that are entailed *a*

*priori* by the facts about the most fundamental ingredients of the world. Recall that, in the Introduction, I provided the following formulation of Jackson's argument supporting the fourth premise of the knowledge argument. Numbers and letters are mixed to highlight the fact that the conclusion is premise four of the knowledge argument. It is merely a coincidence that the conclusion is also the fourth line of the following argument.

#### Jackson's Argument Supporting Premise Four of the Knowledge Argument

- a) If physicalism is true, then the minimal supervenience thesis is true.
- b) If the minimal supervenience thesis is true, then the entailment thesis is true.
- c) If the entry by entailment thesis is true, then there are no facts other than all the physical facts.
- 4. Therefore, if physicalism is true, then there are no facts other than all the physical facts.

The formulation of the entailment thesis in the article co-authored by Jackson and Chalmers focuses on 'microphysical facts'; whereas, when writing alone, Jackson refers to 'physical facts'. In *The Conscious Mind*, Chalmers argues that Jackson's version of the knowledge argument begs the question and also ought to be recast in terms of microphysical knowledge.<sup>171</sup> As I agree with Chalmers that Jackson's ambiguous use of 'physical' leads to question begging (though for different reasons), I will use 'microphysical' when specifically referring to the scientific study of the smallest

constituent parts of the world and I will use ‘physical’ when Jackson’s ambiguity is relevant.

Despite this terminological disagreement, I take it to be an accurate characterization of both Chalmers’ and Jackson’s view that complete microphysical knowledge – along with all that microphysical knowledge entails – is complete knowledge *as far as physicalism is concerned*. They believe that, if physicalism is true, all facts about our world are (in the sense described above) entailed by *P*. Chalmers and (pre-1998) Jackson believe that because Mary knows *P*, but does not know certain facts about chromatic experience, physicalism is false. I believe that it is not possible for Mary to know, by achromatic means, *everything as far as physicalism is concerned*, but I do not doubt that she could learn *P* achromatically. Instead, I deny that *P* (and all it entails) is complete knowledge *as far as physicalism is concerned*. Before I argue that the entailment thesis is not the only possible solution to the placement problem, it will be helpful to specify the sense in which Jackson understands entailment to be an *a priori* matter.

#### B. The Two *A Priori* Aspects of Entailment

There are two *a priori* aspects to Jackson’s entailment thesis. The first is that any given statement expressing a macroscopic truth about the world must follow *a priori* from *P*. The second is that *a priori* conceptual analysis must play a role in determining the reference of macroscopic terms. This second *a priori* aspect involves what has become the controversial two-dimensional interpretation of Kripkean *a posteriori* necessities.<sup>172</sup> It is not my intention to adjudicate issues about how to interpret Kripke;

my objections to the entailment thesis relate primarily to the first sense in which Jackson's entailment thesis makes a claim about what is knowable *a priori*. Nevertheless, the role of conceptual analysis in the entailment thesis must be clarified in order to provide an account of the entailment thesis that can be criticized without attacking a straw version of Jackson's thesis.

### *The Prima Facie Kripkean Objection to the Entailment Thesis*

The entailment thesis presented thus far is the claim that, for physicalism to solve the placement problem, there must be *a priori* entailment directly from the collection of physical facts to any given psychological fact. One might object that since Kripke, nearly all philosophers recognize that statements linking natural kinds (such as 'water is H<sub>2</sub>O') are *a posteriori* necessary truths.<sup>173</sup> On Kripke's view, 'water is H<sub>2</sub>O' is necessary in that it is true in all possible worlds; any worlds without H<sub>2</sub>O are worlds without water, and worlds in which local residents use the word 'water' to refer to something other than H<sub>2</sub>O are worlds in which 'water' does not refer to water. It had to be discovered that 'water is H<sub>2</sub>O'; though necessary, the proposition can only be known *a posteriori*.

*Prima facie*, Jackson's entailment thesis faces a Kripkean objection. Assuming that a completed microphysics does entail psychological facts, the entailment could take the form of natural kind identity statements such as 'seeing red is x pattern of V4 cell activity'. If we accept Kripke's analysis of such statements, and nearly everyone including Jackson does, such statements are necessary, but *a posteriori*. Assuming that the identity of 'seeing red' and 'x pattern of V4 cell activity' is true, it will take empirical investigation to discover the identity. *A priori* analysis of 'seeing red' and 'x pattern of

V4 cell activity' should not be expected to establish an identity any more than *a priori* analysis of 'water' and 'H<sub>2</sub>O' yields an identity. Thus, according to the *prima facie* Kripkean objection, Jackson's entailment thesis should be rejected on the grounds that it sets an unjustifiably high standard for qualitative explanations; no natural kind identity statements are knowable *a priori*.

Relying on a two-dimensional semantic analysis, Jackson factors Kripkean *a posteriori* necessary truths into two parts and thus defends the view that there is an *a priori* part of *a posteriori* necessary truths. Specifically, the kind of *a posteriori* necessary truths that might be candidates for linking a completed microphysics to qualitative facts involve an *a priori* element that plays a part in fixing the reference of the qualitative term, and an *a posteriori* part that connects the qualitative term to the appropriate part of the world.

### *The A Priori Aspect of A Posteriori Necessities*

Consider the following argument:<sup>174</sup>

#### Two-Dimensional Water Argument

- d) The Earth is covered 60% by H<sub>2</sub>O.
- e) H<sub>2</sub>O is the waterish stuff that we find around here.
- f) Water is the waterish stuff that we find around here.
- g) Therefore, the Earth is covered 60% by water.

Notice that premises (d) and (e) are contingent truths, determinable only *a posteriori*. Also notice that the conclusion (g) is a contingent truth. There are, nevertheless, two *a priori* aspects to the argument. First, though the conclusion is itself a contingent truth, it follows *a priori* from the three premises; the argument is deductively valid. More importantly, the conclusion does not follow from (d) and (e) alone. Without (f) the argument is simply invalid. And (f) is asserted to be a necessary *a priori* truth, determinable by conceptual analysis. Specifically, the importance of (f) can be seen in that the unstated sub-conclusion ‘water is H<sub>2</sub>O’ follows from the combination of (e) and (f), along with Leibniz’s Law. Of course, this unstated sub-conclusion is a paradigm example of Kripke’s necessary *a posteriori* truth.

Therefore, the response to the *prima facie* Kripkean objection is “to factor the necessary *a posteriori* statement that water = H<sub>2</sub>O into two parts, a contingent statement about H<sub>2</sub>O that is (assumed to be) derivable from microphysics...[premise e]...and an *a priori* statement that can be justified by conceptual analysis...[premise f]”.<sup>175</sup> So, Jackson agrees with Kripke that statements such as ‘water is H<sub>2</sub>O’ are necessary *a posteriori* truths. Jackson just offers a two part interpretation of *a posteriori* necessities. The first part is a contingent *a posteriori* (physical) claim. The second part is a necessary *a priori* claim (derived from analysis of ‘water’). Only from the combination of these two claims does the necessary *a posteriori* claim follow. Thus, Jackson claims there is an “*a priori* part of the story about the necessary *a posteriori*”.<sup>176</sup>

### *Clarifying the Role of Conceptual Analysis: Two-Dimensional Semantics*

The claim that (f) is an *a priori* and necessary conceptual truth is controversial. In order to understand the sense in which Jackson considers (f) to be an *a priori* truth, it is necessary to clarify the role played by two-dimensional semantics. While there has been a great deal of discussion both regarding two-dimensional semantics in general, and the role it plays in arguments against materialism specifically, the role that two-dimensionalism plays in the entailment thesis is not essential to my objections. Therefore, I will remain neutral as to whether certain terms are two-dimensional and whether one of those dimensions of a term is fixed by *a priori* conceptual analysis.

Central to Jackson's account of two-dimensional semantics is the distinction between the A-intension and the C-intension of a term.<sup>177</sup> The A-intension of a term applies when considering the world that is being contemplated as actual; the C-intension of a term applies when considering a world as counterfactual. Whether these two differ depends on whether the extension of a term differs across possible worlds. To borrow Jackson's example, 'square' is a one-dimensional term: it applies to the same things in the actual world as it does in other possible worlds; the stuff that fills the square role on Twin Earth is no different than what fills the square role on Earth. 'Water', however, is a two-dimensional term: it applies to one kind of thing in the actual world (H<sub>2</sub>O) and something different on Twin Earth (XYZ). "A-extension does not depend on the nature of the actual world" because, with Twin Earth considered as actual, the A-extension of 'water' is to stuff that happens to be XYZ and, with Earth considered as actual, the A-extension of 'water' is to stuff that happens to be H<sub>2</sub>O.<sup>178</sup> The A-extension of 'water' is to whatever fills the role of water in the world that is being considered as actual. In



contrast, when considering a world as counterfactual (that is, when considering the actual world as actual), the C-intension of ‘water’ is whatever is referred to as ‘water’ in the actual world. Thus, in all possible worlds, the C-extension of ‘water’ is to stuff that has been empirically discovered to be H<sub>2</sub>O; the A-extension of ‘water’ varies across possible worlds, being stuff that is H<sub>2</sub>O on Earth and stuff that is XYZ on Twin Earth.

The payoff is that “the sense in which conceptual analysis involves the *a priori* is that it concerns...A-intensions, and accordingly concerns something that does, or does not, obtain independently of how things actually are”.<sup>179</sup> In the Two-Dimensional Water Argument outlined above, “(f) water is the waterish stuff around here” considers the A-intension of ‘water’. That is, the claim that (f) is *a priori* does not assume that the actual world is actual. (f) is true on both Earth and Twin Earth, though it extends to stuff that, as it turns out, has a different chemical make-up in each place. (f) is a necessary truth in that the water role does not change across worlds, but it seems contingent in that the stuff that fills the role does change across worlds.

Thus, as Jackson makes perfectly clear, *a priori* conceptual analysis partially fixes the reference of a term: “we are simply concerned with making explicit what is, and what is not, covered by some term in our language”.<sup>180</sup> Later, he describes statements like ‘water is the waterish stuff we find around here’ as “capturing a fact about reference-fixing”.<sup>181</sup> Specifically, such statements fix the reference of the A-intension of the term. Jackson describes the establishment of this fact about reference fixing as “conceptual analysis” in that it analyzes folk concepts; it asserts what a certain concept refers to when used by ordinary speakers of the language. Though often folk concepts will have to be cleaned up just a bit, modification of the folk concept can not be so

radical that it would no longer be recognizable by the folk. Of course, as Jackson recognizes, it is an *a posteriori* matter that the term ‘water’ is used by folk to refer to the stuff that fills the water role instead of some other term, but the meaning of the term as it is used by the folk is determinable by analyzing the folk use of the concept ‘water’.<sup>182</sup> Jackson also recognizes that folk use of terms is often too vague and confused to serve in arguments like the Two-Dimensional Water Argument. At one point, Jackson describes claims like (f) as involving a “biologically informed folk understanding”.<sup>183</sup> The point is not that science should be limited to folk understandings of terms, but that there are genuine patterns of conceptual competence that are expressible in an ideal language as an articulation of the A-intension of a particular term.

Thus, it is important to note the two senses in which Jackson considers arguments like the Two-Dimensional Water Argument to be *a priori*. One sense is that there is an *a priori* entailment of the conclusion from the premises. Because it is Jackson’s assertion that the entailment involved in the Two-Dimensional Water Argument is the only solution to the placement problem, Jackson believes that the physical facts must, if physicalism is true, play a critical role in entailing any qualitative facts. The second is that the reference of the key term (‘water’) is fixed partially by *a priori* conceptual analysis of folk usage of the term. Thus, Jackson’s defense of the fourth premise of the knowledge argument hinges on a particular interpretation of Kripkean *a posteriori* necessities. On Jackson’s view (before and after 1998), the thought experiment about Mary establishes the falsity of physicalism if Mary is unable to form arguments deriving all genuine psychological facts that are similar to the Two-Dimensional Water Argument. That is, on Jackson’s view, qualitative facts can only be placed in a

physicalist ontology if there are sound arguments about qualitative facts that are of the same form as the Two-Dimensional Water Argument.

### C. Jackson's Knowledge Argument Clarified (Again)

In Chapter Two, I argued that the knowledge argument ought to be seen as a combination of the epistemological claim that qualitative states are informative and the modal claim that it is possible to learn everything as far as physicalism is concerned without qualitative states. Jackson's entailment thesis clarifies the modal claim. Consider the following argument.

#### Two-Dimensional Red Argument

- h) Joseph's V4 cells are firing in pattern  $x$  at time  $t$  (empirical observation, made by Mary of her husband).<sup>184</sup>
- i)  $X$  pattern of V4 cells firing is what fills the seeing red role (contingent fact, derivable from microphysics).
- j) The qualitative experience of seeing red is what fills the seeing red role (*a priori* conceptual truth, based on the A-intension of seeing red).
- k) Therefore, Joseph is having the qualitative experience of seeing red at time  $t$ .<sup>185</sup>

Notice that, in this argument, there is not an *a priori* entailment directly from the complete set of physical facts to any given psychological facts. Rather, there is an *a priori* entailment of qualitative fact (k) from contingent microphysical claim (i) along

with a premise (j), a necessary truth that is arrived at through conceptual analysis. The role of conceptual analysis is to determine the A-intension of qualitative terms, to provide an *a priori* account of what the folk mean when they refer to, for instance, “seeing red”. From this conceptual truth (j), along with empirical observations (h), and truths derivable from microphysics (g) there would be (if physicalism were true) *a priori* entailment of the qualitative fact (k) about Joseph.

The claim of the knowledge argument is that such an argument can not work. Despite the fact that Mary knows h-j inside the black and white study, she learns something when released. Thus, examination of the above summarized argument is a convenient way to evaluate Jackson’s knowledge argument, with the completeness claim – premise four in my formulation – made fully explicit. Nevertheless, Jackson never offers such an argument as an example. His writings that explain the sense in which physicalism is committed to completeness are neutral on the issue of whether psychological facts actually are part of a complete physicalist worldview. Of course, another reason that he never presents an argument similar to the Two-Dimensional Red Argument in his defenses of the entailment thesis is that, by the time of most of that writing, he no longer defends the knowledge argument. In Jackson’s writing, the purpose of the entailment thesis is never (at least explicitly) to defend a premise of the knowledge argument but he would not object to such use of the entailment thesis; the entailment thesis specifies what he has meant by the completeness of the physical facts all along. Such partitioning of issues, though understandable in its own right, detracts from a clear understanding of how the knowledge argument alleges to derive its anti-physicalist conclusion.

So, in what way would defenders of the knowledge argument claim that the above argument fails? First, it is too simple to suggest that Mary cannot know (j) until she leaves the black and white study. After all, (j) is an *a priori* truth about how words are used. Even in the black and white study Mary knows that when people look at a ripe tomato they say “I am seeing red”; she knows under what circumstances folk use qualitative terms. Second, Jackson stipulates that Mary knows (i) inside the black and white room. After all, (i) is derivable from microphysics, and Mary possesses all physical knowledge and everything that can be derived from it. The essence of the knowledge argument is that, though Mary is imprisoned in a black and white room, she knows everything that is relevant for articulating the premises of the above argument. For defenders of the knowledge argument the Two-Dimensional Red Argument fails because there is no way to specify ‘what fills the seeing red role’ in such a way as to bridge the gap between ‘X pattern V4 cells firing’ and ‘seeing red’. Of course, as a matter of basic logic, for the above argument to be valid, ‘what fills the seeing red role’ must be identical in premises (i) and (j). And this is what Jackson claims, at least until his 1998 conversion to physicalism, can not be done. The specification of ‘what fills the seeing red role’ in (i) must be in the language of the physical sciences. However, the specification of ‘what fills the seeing red’ in (j) can not be in the language of the physical sciences, because qualitative terms such as ‘seeing red’ are used by the folk in a way that excludes such scientific descriptions. The qualitative aspect of them is essential to their application, and any description of their application in the language of physical sciences will necessarily leave something out.<sup>186</sup>

Put more colorfully, imagine that one day Mary is given a research partner. Ronald is a fictional super-phenomenologist, able to describe everything there is to be described about the content of qualitative states. Mary is given the task of working out premise (i) in luxurious detail; Ronald is given the task of working out premise (j) in luxurious detail. Even given an infinite amount of time, defenders of the knowledge argument assert that Mary and Ronald will not be able to agree on a specification of 'what fills the seeing red role', because they will forever use incompatible language for describing 'what fills the seeing red role'. Therefore, the qualitative fact expressed in the conclusion of the argument is not entailed by the combination of Mary's exhaustive physical facts and Ronald's exhaustive understanding of the A-intension of 'seeing red', and there is no place for qualitative facts in a physicalist ontology.

## II. The Entailment Thesis, Qualitative Facts, and the Placement Problem

I agree with pre-1998 Jackson that qualitative properties can not be eliminated. I further agree that physicalism faces the placement problem regarding qualitative properties. Physicalists, in order to offer a complete metaphysical view, must locate qualitative properties among the physical properties of the world. However, I will argue that the entailment thesis is not how the placement problem gets solved for biological facts, and therefore should not be expected to be how the placement problem should be solved for qualitative facts. First, I will argue that the entailment thesis can not effectively respond to the critiques I made of the knowledge argument, because those critiques apply equally to the entailment thesis. Second, I will argue that there is nothing about the entailment thesis that supports the claim that biological facts reduce to physical

facts but qualitative facts do not. The more stringent standard he sets for the reduction of qualitative facts, the less plausible will be the claim that biological facts meet the standard. The looser he sets the standards for reduction of biological facts, the less plausible will be the claim that qualitative facts can not meet the standard.<sup>187</sup>

#### A. The Entailment Thesis Does Not Address Critiques of Chapters Two and Three

In Chapters Two and Three I argued that Jackson's knowledge argument conflates metaphysical and epistemological issues and that, as a result, both the metaphysical and epistemological significance of the thought experiment about Mary have been overstated. In Chapter Two, I highlighted Jackson's ambiguity about the meaning of 'physical' and argued that, as a result, his stipulation that Mary knew all of the physical facts in the black and white room begged the question regarding the physicalist credentials of qualitative facts. In this section, I will argue that the entailment thesis also conflates metaphysical and epistemological issues, and that Jackson's persistent failure to define 'physical' continues to lead to question begging.

#### *The Entailment Thesis Conflates Metaphysics and Epistemology*

I agree with Jackson that the minimal supervenience thesis ("any world that is a minimal physical duplicate of our world is a duplicate *simpliciter* of our world") is a commitment of physicalism. Additionally, Jackson believes that this supervenience thesis is true only if the entailment thesis is also true ("any psychological fact about our world is entailed by the physical nature of our world").<sup>188</sup> Of course, 'fact' is most naturally taken in a purely metaphysical sense: the facts are the facts whether we have

epistemic access to them or not. I also agree with Jackson that a purely metaphysical interpretation of the entailment thesis is a commitment of physicalism.<sup>189</sup> However, Jackson also writes that, “psychological facts have a place in the physicalists’ world view if and only if they are entailed by some true, purely physical statement”.<sup>190</sup> The primary difference between the above two statements is in the final clause that names what entails the psychological facts. In the first, it is the physical *nature of the world* (a metaphysical notion) that must entail psychological facts; in the second it is physical *statements* (a linguistic notion) that must entail psychological facts. The replacement of metaphysical terms with linguistic ones is completed when he goes on to claim that the physical “story” about the world must entail the psychological “story”. It is also important to distinguish the kind of linguistic entailment that Jackson describes from an epistemological understanding of ‘entailment’. After all, it can be natural to interpret ‘entailment’ epistemologically: *the story we can tell* about the physical nature of our world entails *the story we can tell* about the psychological nature of our world. This latter, epistemological, reading of ‘entailment’ is of an intertheoretic explanatory relationship in which *everything that science could discover* about psychological facts is reductively explained by *everything that science could discover* about the physical facts.

In short, it is necessary to distinguish between three senses of ‘entailment’:  
*metaphysical entailment, linguistic entailment, and epistemological entailment.*

*Metaphysical entailment* is a relationship between facts, such that one set of facts makes true another set of facts; one set of facts is true in virtue of another set of facts. If the



psychological facts are metaphysically entailed by the microphysical facts, then a minimal physical duplicate of our world is also a psychological duplicate of our world.

*Linguist entailment* is a relationship between sentences in some idealized language. If the psychological facts are linguistically entailed by the microphysical facts, then the collection of microphysical statements expressible (by a God) in an idealized language must entail the psychological statements that can be expressed in such an idealized language.

*Epistemological entailment* is a relationship between actually discoverable scientific theories; it is a matter of reductive explanation. If the psychological facts are epistemologically entailed by the microphysical facts, then it is possible (in principle) for human beings to make discoveries by means of the natural sciences from which, along with bridge laws, the qualitative facts about the world follow deductively.

I agree that, because physicalism is committed to the minimal supervenience thesis, physicalism is committed to a metaphysical entailment thesis. I do not think that physicalism is committed to either a linguistic or an epistemological entailment thesis. Jackson's pre-1998 anti-physicalist conclusions (and his post-1998 *a priori* physicalist conclusions) rest on a conflation of these three senses of entailment.

Given that Jackson describes Mary as an "actual person" and writes about what she *knows*, discussion of the knowledge argument is most naturally cast in terms of the epistemological sense of 'entailment'. That is, with the thought experiment about Mary,

Jackson argues (plausibly) that psychological facts are not epistemologically entailed by the microphysical facts. Even a scientifically omniscient knower could not discover theories in microphysics that explain a possible theory of qualitative psychology. Instead of asserting that physicalism is committed to epistemological entailment, however, Jackson asserts (in his writings on the entailment thesis) that physicalism is committed to linguistic entailment. If the entailment thesis is intended to clarify the fourth premise of the knowledge argument, then it seems that the knowledge argument equivocates on three different senses of a complete set of physical facts, not just two. From the claim that qualitative facts are not epistemologically entailed by microphysical facts (roughly, premise three of the knowledge argument), and the claim that physicalism is committed to the linguistic entailment of the qualitative facts from the microphysical facts (roughly, premise four of the knowledge argument), Jackson (pre-1998) concludes that qualitative facts are not metaphysically entailed by microphysical facts (in that he believes the existence of qualitative facts demonstrates the incompleteness of physicalism). What Jackson is missing is an argument claiming that because physicalism is committed to metaphysical entailment, physicalism is also committed to linguistic entailment. Also missing is an argument claiming that if epistemological entailment fails then linguistic entailment also fails. I do not think that the missing parts of Jackson's arguments can be filled in.

For the sake of argument, I will grant that any set of facts  $x$  that epistemologically entails another set of facts  $y$  will also linguistically, and thereby metaphysically, also entail  $y$ . However, just because the physical facts metaphysically entail the psychological facts does not mean that the physical facts linguistically entail

the psychological facts. A linguistic story is a representation of the facts of the world, created by some intellect or community of intellects. Even if one set of facts about the world entails another set of facts about the world, there is no reason to believe that *the story that can be told* about one set of facts entails *the story that can be told* about another set of facts. Physicalists are only committed to linguistic entailment if they believe, in addition to physicalism, that an idealized language can express all of the truths about the world. It seems that even an idealized language is a representation of the world, and it is perfectly plausible to hold that qualitative facts are determined by microphysical facts in ways that are not expressible in any language, even by a God. Perhaps some truths about the physical world are absolutely ineffable.

Jackson might respond that, by definition, an idealized language provides the means (at least to God) for expressing every truth that there is about the world, and that my imagination is being unduly limited by the expressive power of human language. If so, I would respond that he has only collapsed linguistic entailment and metaphysical entailment, making it even more clear that there is a gulf between the complete microphysical knowledge attributed to Mary inside the black and white room and the sense in which physicalism is committed to completeness. If Jackson limits his argument to the claim that physicalism is committed to such a robust version of the linguistic entailment of qualitative facts by the microphysical facts, then he needs to show that it is possible for someone locked in a black and white room to know all the physical facts in the sense of 'all the physical facts' that is relevant to such a robust version of linguistic entailment. Yet, if he were to recast the thought experiment about Mary in terms of linguistic entailment, the intuition that Mary would learn something on her release is

weakened; who can guess what a God might be able to figure out? In Chapter Two, I claimed that certain qualitative facts (the ones about chromatic vision) can only be known chromatically, and that such facts would also be true in a minimal physical duplicate world. The position I defended in Chapter Two amounts to the claim that qualitative facts are metaphysically, but not epistemologically, entailed by the microphysical facts. Jackson's introduction of linguistic entailment does not bridge the gap between the claim that physicalism is committed to a metaphysical entailment thesis and the claim that qualitative facts are not epistemologically entailed by microphysical facts.

In addition to denying that all examples of metaphysical entailment are examples of linguistic entailment, I also deny that all examples of the failure of epistemological entailment are examples of the failure of linguistic entailment. Science, even completed science, is a deeply human enterprise in which theories emerge that are constrained by expressive capacities of human language and shaped by the structures of human perception. Scientific explanations provide human beings with a kind of intellectual satisfaction that would be lacking in Jackson and Chalmers' account of linguistic entailment. The conjunction of all microphysical truths about the world (what Jackson and Chalmers' designate as *P*) may linguistically entail ideal language accounts of biological and chemical facts, but *P* would provide human beings no intellectually satisfying explanations of biology and chemistry. No human being could understand *P*, because *P* would be too long and complicated. Even some part of *P*, such as the conjunction of all the microphysical truths about a turtle, could never provide any human

being an intellectually satisfying answer about why the turtle climbed into its shell when the shark came by, or what the shell is made of.

Thus, the clarification of the knowledge argument offered by the entailment thesis highlights, rather than mitigates, Jackson's conflation of metaphysical and epistemological issues. Assume that Ronald the super-phenomenologist and Mary the super-neuroscientist will never agree on how to characterize 'what fills the seeing red role'. This is a failure of one way of seeking knowledge to connect with another way of seeking knowledge: a failure of one "story" to connect to another "story". Yet, the purpose of the entailment thesis is to explain the sense in which physicalism is committed to the completeness of the physical facts and, when seen as clarification of the fourth premise of the knowledge argument, to connect the epistemological thought experiment about Mary to the knowledge argument's overtly metaphysical conclusion. In the end, Jackson's account of linguistic entailment provides no reason to deny that chromatically accessed facts about subjective experience are made true by the microphysical nature of the world.

### *Jackson is Still Begging the Question*

In Chapter Two, I argued that Jackson's refusal to define 'physical' leaves him with no non-question begging justification for refusing the label 'physical' to qualitative facts. When defending the entailment thesis, Jackson continues to use 'physical' ambiguously even after his 1998 renunciation of the epistemological claim of the knowledge argument. In 1994 Jackson writes that by 'physical' or 'physical property' he just means what is usually meant, and that he considers the controversies regarding the

meaning of these terms to be irrelevant to his argument.<sup>191</sup> In 1998 Jackson suggests that problems in defining 'physical' are "more apparent than real".<sup>192</sup> In 2001, Chalmers and Jackson state "we will not engage the issue of what counts as 'physics'".<sup>193</sup> The result of Jackson's ambiguity about the meaning of 'physical' is that the linguistic entailment thesis fails to provide a non-question begging reason to deny that qualitative facts deserve the label 'physical facts' in the sense that is relevant to the completeness of physicalism.

In his writings on the entailment thesis, Jackson contends that his vagueness allows him to avoid two controversies that have no bearing on the entailment thesis: 1) why chemistry, physics, and biology are privileged over psychology or economics, and 2) how committed physicalism is to the current state of these sciences.<sup>194</sup> Whether these two controversies have a bearing on the argument depends on which argument we are talking about. For the argument that conceptual analysis is an important metaphysical tool (the stated purpose of his writings on conceptual analysis), I agree that the precise definition of 'physical' is irrelevant. However, in understanding physicalism's claim to completeness, the controversies over the privileging of physics, chemistry, and biology over psychology or economics, and the extent to which physicalism is committed to the current state of these sciences matter a great deal. Regarding the controversy over physicalism's commitment to the current state of the natural sciences, consider the Churchland-Dennett position that a completed physics might entail a great many things that all of us could never have imagined. By focusing on what is knowable in principle, Jackson seems to suggest that physicalism is not committed to the current state of the natural sciences. Yet, Jackson asks us to assume that a completed science will provide

many more propositions of the same sort that the natural sciences currently provide, that it is a fair bet we have got the basics right.<sup>195</sup> He seems to want it both ways: he recognizes that scientific knowledge acquirable in principle is much broader than science as we understand it, yet he still runs his thought experiments on the assumption that the scientific knowledge acquirable in principle profoundly resembles the kind of stuff that is found in the current state of the natural sciences.

Regarding the controversy about privileging physics, chemistry, and biology, it is particularly instructive that psychology is excluded from what he calls the “privileged story”. Jackson simply defines psychology out of *physical* and gives no other reason than, apparently, convention among physicalists. Without a clear definition of ‘physical’ it is impossible to determine whether the privileging of physics, chemistry, and biology is justified in the context of metaphysical claims regarding physicalism’s claim to completeness. To exclude psychological investigations from the physicalist story on the grounds that it is left out of what is usually meant by ‘physical’ or ‘physical information’, and to do so while arguing for a certain view of the sense in which physicalism is committed to completeness, is to beg the question. After all, the question of what counts as part of the ‘physical’ story will profoundly impact how plausible the linguistic entailment thesis seems. The more broadly Jackson conceives of the physical base of the linguistic entailment, the weaker is the claim that the physical story does not entail the qualitative story. The more narrowly Jackson conceives of the physical base of the linguistic entailment, the weaker is the claim that that physicalists must hold that the physical facts linguistically entail the qualitative facts. Instead of explaining the criteria for including certain facts (chemical and biological) in the physical story while excluding

others (psychological), Jackson evades the question. In the end, it is clear neither what set of facts is to do the entailing nor why there can be no chromatically accessed facts among the entailing facts. Besides, most psychologists would take offense at the suggestion that they are investigating something non-physical. And the psychological facts do not need to be linguistically entailed by the physical facts if they are among the physical facts. All in all, it seems rash to exclude all psychological facts, even qualitative facts, from the domain of physical facts – and to do so from the armchair.

### *An Eliminativist Objection*

Most of this chapter focuses on entailment and remains neutral regarding Jackson's two-dimensional semantics. This brief section is the only exception. Jackson claims that the *a priori* part of Kripkean *a posteriori* necessities involves fixing the reference of folk concepts. Specifically, if physicalism is true, there must (in principle) exist an account of the meaning of folk psychological concepts that is identical (enough) to an account of the related scientific concepts. For instance, 'seeing red' (or 'memory', 'belief') must be definable in a single way that accurately reflects both folk and neuroscientific usage of the phrase. Jackson's account of the role of conceptual analysis in determining the A-intension of two-dimensional terms (a role he summarized as providing "biologically informed folk concepts") explains the role of folk concepts. The structure of arguments such as the Two-Dimensional Red Argument formally require that the account of the folk concept must be identical enough to the account provided by microphysics ('the seeing red role' must be univocal in (i) and (j)). Given the lack of



philosophical and scientific sophistication among ordinary folk this is an incredible claim.

Of course, as Jackson recognizes, physicalists can choose to eliminate psychological claims instead of placing them (in the sense of the placement problem). However, it is one thing to say that qualitative states are not real and quite another to say that *the folk conception of* qualitative states are entirely mistaken. Call this a distinction between ontological eliminativism and conceptual eliminativism. Ontological eliminativism is the claim that the folk conception of a qualitative state is a conception of something that is not real, such as ‘demons’ in the demonic possession theory of seizure disorders. Conceptual eliminativism is the narrower claim that the folk conception of a qualitative state is a misunderstanding (even a radical misunderstanding) of something that is real, such as the understanding of ‘seizures’ in the demonic possession theory of seizure disorders. I agree that ontological eliminativism regarding qualitative mental states is untenable; there is something that it is distinctly like to hear a telephone ring, to feel velvet, to taste cotton candy, or to see red.<sup>196</sup> However, conceptual elimination of some qualitative terms is not only tenable, it is supported by a great deal of neuropsychological research.<sup>197</sup> For instance, whereas the folk consider memory to be a unitary phenomenon, scientists have learned from patients such as HM that there is an important distinction to be drawn between different kinds of memory; similar things can be said for attention. That folk concepts about memory have followed the research (folk speak of long term and short term memory all the time now) suggests that even the folk recognize that folk conceptions can be misleading. As explained at length in Chapter Three, Patricia Churchland’s eliminativism is the claim that folk psychological concepts

might need to be abandoned in favor of scientific concepts, not the claim that folk psychological concepts refer to illusions. It is more than plausible to accept conceptual eliminativism without accepting ontological eliminativism.

In granting that ordinary folk refer to *something* when they use the phrase ‘seeing red’ one need not claim that the A-intension of ‘seeing red’ refers to the same genuine features of the world that a scientific account might discover. Physicalists can consistently contend that folk usage of qualitative terms must be abandoned in order to explain qualitative states, while also holding that qualitative states are genuine physical phenomenon, and that statements in a future scientific language might epistemologically (let alone linguistically) entail qualitative facts which are themselves expressed in the language of a future objective phenomenology.<sup>198</sup> Kripkean *a posteriori* necessities may, or may not, involve an *a priori* part that connects ordinary language to scientific language. To suggest that there could never be such an *a priori* part for *a posteriori* necessities regarding qualitative states (as the knowledge argument does when the entailment thesis is presumed to be an explication of premise four) is not an objection to physicalist accounts of qualitative states.

## B. The Entailment Thesis Fails to Drive a Wedge between Qualitative Facts and Other Facts

Taking the linguistic entailment thesis as an explication of the fourth premise of the knowledge argument, the knowledge argument suggests that qualitative facts are not linguistically entailed by the microphysical facts. Jackson’s account of Mary’s pre-release knowledge in terms of a completed physics, chemistry, and biology indicates that

Jackson believes chemical and biological sciences are linguistically entailed by the microphysical facts. The knowledge argument only works if qualitative facts are unlike biological facts in that they are not linguistically entailed by the microphysical facts. Metaphorically, Jackson must drive a wedge between qualitative facts on the one hand and chemical and biological facts on the other hand.<sup>199</sup> In this section, I will argue that the entailment fails to drive such a wedge because whatever one might say about biological facts and entailment, one can also say the same thing about qualitative facts. Jackson and Chalmers provide no compelling reason to believe that microphysical facts linguistically entail facts about life and water but not facts about what it is like to see red. On my view, physicalists should solve the placement problem for qualitative facts in the same way as they have for biological facts.

*Biological Cases of Reductive Explanation do not Provide Examples of Epistemological Entailment*

Ned Block and Robert Stalnaker argue that Jackson and Chalmers only defend the narrow claim that “*if a conceptual analysis of a certain kind were always available, then we could use these conceptual analyses to account for the necessary *a posteriori* truths of reductive explanation*”.<sup>200</sup> They grant that epistemological entailment (Nagel-reduction) establishes a sufficient condition for reductive explanation, but that reductive explanation can occur in the absence of epistemological entailment. As Block and Stalnaker point out, “not a single example of an analysis of a non-microphysical term in terms of microphysics is given either by Jackson or Chalmers”.<sup>201</sup> In explaining and defending their two-dimensional understanding of the linguistic entailment thesis,

Jackson and Chalmers provide purported examples of the linguistic entailment of microphysical facts such as the Two-Dimensional Water Argument outlined above. If we consider paradigm cases of reductive explanation in the biological sciences we find (according to Block and Stalnaker) that such epistemological entailments are lacking. For instance, few would deny that scientists have provided a reductive explanation of life in terms of natural biological processes. “One might try to analyze life *a priori* in terms of reproduction, locomotion, digestion, excretion, respiration, and the like, and then give further analysis of these terms, eventually grounding the functions in microphysical terms.”<sup>202</sup> But Block and Stalnaker claim that such an attempt would be hopeless. Their view is that statements about reproduction, locomotion, and so on, do not epistemologically entail statements about life, even though it is widely accepted that life is reductively explained (and certainly placed within a physicalist ontology) by precisely such statements.

Block and Stalnaker argue that the properties relevant to the reductive explanation of life (reproduction, locomotion, digestion, excretion, respiration) are not necessary for being alive. We can imagine discovering something that we consider to be living that possesses none of these attributes. Block and Stalnaker imagine something that is immortal (does not need to reproduce), is tree-like (does not locomote), takes from the soil only what it needs (does not excrete), and needs nothing from the air (does not respire). While such an organism is quite difficult to imagine, there is no conceptual reason to deny that such a thing is alive. Block and Stalnaker also argue that these criteria are not sufficient for considering something alive. “A moving van locomotes, processes fuel and oxygen, and excretes waste gases. If one adds a miniature moving van

factory in the rear, it reproduces. Add a TV camera, a computer, and a sophisticated self-guiding computer program, and the whole system could be made to have more sophistication, on many measures, than lots of living creatures.”<sup>203</sup> The point is not that it is impossible to consider such a contraption to be living (though I think we should not); the point is that there is no conceptual obligation to consider it living. Even widely accepted cases of reductive explanation do not offer examples of epistemological entailment. Epistemological entailment is not how the placement problem gets solved for biological facts.

Jackson and Chalmers respond that the entailment thesis only requires that *a priori* entailment of the facts about life from the physical facts be available in principle to an idealized observer (to a God, as it is sometimes put). In essence, they respond by pointing out that they intend linguistic entailment whereas Block and Stalnaker have only argued that there is no epistemological entailment of biological facts from microphysical facts. It is (I think) Jackson and Chalmers’ view that, though biological facts are not epistemologically entailed by the microphysical facts (human science can never provide a Nagel-style reduction of biology to microphysics), biological facts are nevertheless linguistically entailed by the microphysical facts (biological facts follow deductively from God’s account of the microphysical facts). As Jackson and Chalmers put it, “Block and Stalnaker’s discussion does not engage the first-order issue of whether the *a priori* entailments in question exist”.<sup>204</sup> Recall that, by an ‘epistemological entailment’ thesis, I mean the thesis that it is possible in principle for human beings to make discoveries by means of the natural sciences that can be expressed in sentences that entail macroscopic facts (such as biological and qualitative facts). Block and Stalnaker

only argue that a human science of microphysics can not epistemologically entail biological facts; they do not demonstrate that biological facts are not entailed by God's account of the microphysical facts.

But (echoing an argument I made in Chapter Three) there are no grounds for making claims that  $x$  would be known by an ideal observer even though  $x$  can not, even in principle, be known by any actual observer. If anything, the failure of  $x$  to be epistemologically entailed by microphysics creates substantial *prima facie* doubt that  $x$  is nevertheless linguistically entailed by microphysics. Jackson and Chalmers seem willing to grant that statements humans could actually construct about biological facts will never be entailed by the collection of statements humans could actually construct about microphysical facts. As a result, it seems to me, they have ruled out the existence of any citable evidence for their assertion that microphysical statements that could be constructed in an idealized language entail the ideal language statements about biology. Though epistemological and linguistic entailments are importantly different, to acknowledge a failure of epistemological entailment in any particular case is to acknowledge that there is no available evidence that such a case provides an example of linguistic entailment. If biological facts are not linguistically entailed by microphysical facts then, apparently, linguistic entailment is not the only solution to the placement problem (as biological facts clearly have a place in a physicalist ontology). On the other hand, any defense of the claim that biological statements are linguistically, though not epistemologically, entailed by microphysical statements could equally be applied to qualitative facts. That is, to the extent that a case can be made that the biological facts

are linguistically entailed by the microphysical physical facts, the case against the linguistic entailment of the qualitative facts by the microphysical facts is weakened.

### *The Wedge Claim, Specified*

To understand and evaluate Chalmers' and (pre-1998) Jackson's claim that there is a wedge to be driven between qualitative facts and other kinds of facts, it is most helpful to compare three different two-dimensional arguments: one qualitative, one biological, and one microphysical. I have already presented the qualitative and microphysical examples. By way of reminder, the following argument about water is Jackson's paradigm example of the kind of relationship between microphysical facts and macroscopic facts that is (according to Jackson) necessary for physicalists to solve the *placement problem* and place macroscopic facts in a physicalist ontology.

### Two-Dimensional Water Argument

- d) The Earth is covered 60% by H<sub>2</sub>O. (empirical observation)
- e) H<sub>2</sub>O is the waterish stuff that we find around here. (contingent and *a posteriori*; derivable from microphysics)
- f) Water is the waterish stuff that we find around here. (determined by *a priori* conceptual analysis, the A-intension of 'water')
- g) Therefore, the Earth is covered 60% by water.

Given that the Two-Dimensional Water Argument is the form of argument that it must be possible (for God) to construct in order to place water in a physicalist ontology, and

that the knowledge argument purports to demonstrate that qualitative facts do not meet the necessary conditions for being placed in a physicalist ontology, I earlier suggested that the knowledge argument can be expressed in terms of the inability of even an idealized observer to specify the Two-Dimensional Red Argument in such a way that the argument is sound. By way of reminder:

#### Two-Dimensional Red Argument

- h) Joseph's V4 cells are firing in pattern  $x$  at time  $t$  (empirical observation, made by Mary of her husband).
- i)  $X$  pattern of V4 cells firing is what fills the seeing red role (contingent fact, derivable from microphysics).
- j) The qualia of seeing red is what fills the seeing red role (*a priori* conceptual truth, based on the A-intension of seeing red).
- k) Therefore, Joseph is having the qualitative experience of seeing red at time  $t$ .

To explain the sense in which defenders of the knowledge argument would claim that this Two-Dimensional Red Argument fails, I earlier imagined that Mary is given a research partner named Ronald, a super-phenomenologist whose job it is to specify (j) in luxurious detail. Defenders of the knowledge argument claim that no matter how much conceptual analysis Ronald does of the A-intension of what it is like to see red, and no matter how much microphysics Mary knows, the two of them will never be able to agree on a unifying characterization of 'what fills the seeing red role'. Whatever Mary proposes will be rejected by Ronald because it will not capture the subjective experience



of what it is like to see red; whatever Ronald proposes as an informed folk understanding of what it is like to see red will be just so much gibberish to black and white Mary.

Chalmers and (pre-1998) Jackson argue that there is a wedge between the above two arguments, such that an ideal observer could construct highly specific statements so that the argument about water is sound, but could not construct highly specific statements so that the argument about seeing red is sound. Since arguments of this form are alleged to be the only solution to the *placement problem*, Jackson concludes that ‘what it is like to see red’ has no place in physicalist ontology. In order for me to explain why I think the case is not made that such a wedge exists, we need to also consider a third two-dimensional argument:

#### Two-Dimensional Life Argument

- l) Ankata<sup>205</sup> (my pet lizard) respire, digests, locomotes etc. (empirical observation).
- m) Respiring, digesting, locomoting etc. is *what fills the such and such role* (microphysical truth).
- n) Being a living creature is *what fills the such and such role* (conceptual truth).
- o) Therefore, Ankata is a living creature.

Block and Stalnaker effectively argue that the Two-Dimensional Life Argument does not offer an example of epistemological entailment; it is common ground that the Two-Dimensional Life Argument does offer an example of metaphysical entailment (no one argues that biological facts fail to supervene on the microphysical facts). On Jackson and

Chalmers' view, the Two-Dimensional Water Argument offers an example of metaphysical, linguistic, and epistemological entailment. Defenders of the knowledge argument must hold that the Two-Dimensional Red Argument does not offer an example of metaphysical, linguistic, or epistemological entailment. However, the problem that advocates of the knowledge argument have with the Two-Dimensional Red Argument also seem to apply to the Two-Dimensional Life Argument. That is, it seems impossible to specify the term 'what fills the such and such role' in a unified way such that (m) is a real microphysical truth and that (n) accurately reflects what ordinary folk mean 'being a living creature'. We might imagine parallels to Mary and Ronald, one whose job it is to work out (m) in luxurious detail and one whose job it is to work out (n) in luxurious detail; I see no reason to believe that such parallels to Mary and Ronald would ever agree on a single characterization 'what fills the such and such role'. First of all, respiring, digesting, locomotion etc. are so different from one another that there doesn't seem to be anything coherent that could follow *from the conjunction of them* as an *a posteriori* microphysical truth. Secondly, specifying (n) seems equally hopeless. What even the biologically informed folk mean by 'being a living creature' is likely not coherent enough to be helpful in a reductive explanation. If Jackson responds by claiming that, by 'being a living creature' biologically informed folk mean the conjunction of the properties listed in (l) and (m), then not only does he beg the question about the entailment of the A-intension of life from microphysics, he also opens another argumentative strategy to those wishing to place qualitative facts in a physicalist ontology: when the folk are sufficiently biologically informed, their conception of seeing

red will be such that facts about seeing red might be reductively explained by microphysics.

### *The Wedge Claim Rejected*

Clearly, Jackson would say that biological facts are metaphysically and linguistically entailed by the microphysical facts, but he faces a dilemma regarding epistemological entailment of biological facts. It would be reasonable to argue that biological facts are not epistemologically entailed by microphysical facts, because no amount of human science could produce theories in biology and microphysics suitable for a Nagel-reduction of qualitative psychology to microphysics. It would also be reasonable to argue that such a Nagel-reduction is available to human science in principle, the science just hasn't been developed yet. As I already mentioned, Jackson and Chalmers seem to opt for the first horn of the dilemma (conceding the point to Block and Stalnaker). Nevertheless, whichever horn of the dilemma they choose, a case can be made that qualitative facts can be placed within physicalist ontology in the same way that biological facts are.

Assume that Jackson holds, as he seems to, that only the Two-Dimensional Water Argument is a genuine example of epistemological entailment; neither the Two-Dimensional Red Argument nor the Two-Dimensional Life Arguments provide genuine examples of epistemological reduction. If such an assumption is correct, then Jackson has no reason to deny of qualitative facts that which he has asserted of biological facts: they are linguistically (and thus metaphysically) but not epistemologically entailed by microphysical facts. Put differently, if the failure of the biological facts to be

epistemologically entailed by the microphysical facts does not prevent biological facts from being *placed* within a physicalist ontology, then there is no reason why the failure of qualitative facts to be epistemologically entailed by the microphysical facts should have any ontological significance. After all, as mentioned earlier, Jackson's description of Mary as *knowing* a completed *physical science* suggests that the thought experiment about Mary can only establish the failure of qualitative facts to be epistemologically entailed by the microphysical facts. Since the failure of epistemological entailment for biological facts has no metaphysical significance, the failure of epistemological entailment for qualitative facts need not either.

Assume, counterfactually (I think), that Jackson holds that the two-dimensional life example does provide a genuine example of epistemological entailment. That is, Jackson might argue that it is possible in principle to develop a human science that entails the facts about life – even though it has not (yet) been done in practice. If so, then the fact that current microphysical theory does not reductively explain the qualitative facts can not be taken as evidence that a completed microphysical theory would not epistemologically entail the qualitative facts. Put differently, if it is merely the current state of science that keeps us from understanding the epistemological entailment of biological facts from microphysical facts, then there is no reason to conclude that the failure of current science to reductively explain qualitative facts has any ontological significance.

Either way, whatever Jackson says about the soundness of the Two-Dimensional Life Argument, one might say the same thing about the Two-Dimensional Red Argument. To argue that a wedge can be driven between the Two-Dimensional Water

Argument and the Two-Dimensional Red Argument is not enough. Unless he can also drive a wedge between the Two-Dimensional Red Argument and the Two-Dimensional Life argument, he has no justification for the claim that physicalists can solve the placement problem regarding biological facts but not qualitative facts. Without a wedge between qualitative and biological facts, my recognition of the existence of qualitative facts is no more of a threat to the completeness of my physicalist ontology than is my recognition of biological facts. As Valerie Hardcastle puts it, “all are on a par”.<sup>206</sup>

## Conclusion

All of which takes us back to contemplation of Mary, the achromatic knower of all the facts that are countenanced by physicalist ontology. Even assuming that chromatic experiences are the only way to learn qualitative facts, there is no reason to suggest that qualitative facts can not have a place in physicalist ontology. On any available account, qualitative facts have as much physicalist credibility as biological facts do. Of course, by rejecting Jackson’s entailment thesis I have only argued that linguistic entailment is not the only way to solve the placement problem. In the Conclusion and Appendix, I offer a speculative account of the placement of qualitative facts in a physicalist worldview.

---

## Notes to Chapter Four

168 Jackson and Chalmers, “Conceptual Analysis and Reductive Explanation,” 316.

169 However, the conjunction of *P* and *T* “does not imply any indexical truths. Therefore, *I* is stipulated to be locating information, and the conjunction of *P*, *T*, and *I* “will imply all indexical truths”. To keep the issue of the entailment of qualitative facts from physical facts and indexical truths to the side, *Q* is stipulated to be “the conjunction of all phenomenal truths”. Thus, stated in perhaps its most explicit version, the entailment thesis is that *PQTI* entails *M*. See Jackson and Chalmers, “Conceptual Analysis and Reductive Explanation, 316-319.

170 Jackson and Chalmers, “Conceptual Analysis and Reductive Explanation,” 316.

171 Chalmers, “Phenomenal Concepts and Knowledge Argument,” 285.

172 Some might object to my emphasis on *a prioricity*, arguing that the emphasis should not be on whether there is something *a priori* about the entailment, but rather than on the fact that there is an entailment at all.

---

I respond only that Jackson emphasized *a prioricity* because he thinks that, if physicalism is true then the physical facts (along with conceptual truths) entail all the other facts *without further assistance*.

173 Kripke, *Naming and Necessity*, 123.

174 For this formulation I rely on Block and Stalnaker, "Conceptual Analysis and Explanatory Gap," 12. Similar arguments can be found in Jackson, "Armchair Metaphysics," 34; and Jackson, *Metaphysics to Ethics* 59, 74. For essentially the same argument, see Jackson, *From Metaphysics to Ethics*, 82. I chose to rely on Block and Stalnaker's formulation rather than one of these formulations by Jackson because Jackson chose to introduce two-dimensional semantics before the analysis of necessary *a posteriori* truths, and his formulations therefore rely on the distinction between A-intensions and C-intensions which I found it more helpful to introduce later.

175 Block and Stalnaker, "Conceptual Analysis and Explanatory Gap," 12.

176 Jackson, "Armchair Metaphysics," 37.

177 The same distinction is made by a number of other philosophers with different terminology. For instance, Chalmers's, *The Conscious Mind* refers to the primary and secondary intension of a term. Chalmers, "Epistemic Two-Dimensional Semantics," 160 refers to the 1-intension and 2-intension. Since 'A' stands for 'actual' and 'C' stands for 'counter-factual' in Jackson's formulation, I consider it to be the most useful terminology. For other terminology used, as well as lengthy discussion of differences between various approaches to the two-dimensional distinction, see Chalmers "Epistemic Two-Dimensional Semantics," 163-5.

178 Jackson, *Metaphysics to Ethics*, 50.

179 Ibid, 51.

180 Ibid, 51.

181 Ibid, 59.

182 Ibid, 47.

183 Ibid, 64.

184 Of course, in order to make such an observation Mary will need some very sophisticated scanning equipment.

185 For a similar argument, based on experiencing pain rather than seeing red, see Block and Stalnaker, "Conceptual Analysis and Explanatory Gap," 26. They use their argument about pain to support the claim that an argument of this structure is just as plausible when the terms refer to qualitative mental states as it is when the terms are more overtly physical. They argue that the conceptual claim in both arguments I outlined (premises c and g) "stand or fall together". Though I think they are probably right about this claim, the claim seems to miss the point of the knowledge argument. The Mary thought experiment shows that the relevant disanalogy between the two arguments is in not in the asserted conceptual truth but in the connection between the conceptual claim and the microphysical claim. After all, in the black and white study Mary will know everything there is know about the conditions under which people use specific qualitative terms.

186 One might suggest that, in 'what fills the seeing red role', 'the seeing red role' might be spelled out functionally, thus providing a bridge between (i) and (j). That is, (i) might be reformulated 'V4 cells firing is what makes people say that they see red' and (j) might be reformulated 'seeing red is what makes people say that they see red'. But it is not hard to see that, on such a formulation, (j) is vacuous. People mean more by 'seeing red' than merely the conditions under which they say they are seeing red. They mean the rich qualitative texture of redness. And it is exactly this qualitative texture that Jackson claims can not – even in principle – be described in the only language that is available for specifying (i) (the language of the physical sciences).

187 Block and Stalnaker, "Conceptual Analysis and Explanatory Gap," 27.

188 Jackson, "Armchair Metaphysics," 31.

189 For the sake of clarity, I will limit myself to the metaphysical sense of 'fact'. I point out that 'fact' can be taken in either sense merely to highlight the subtlety of Jackson's shift from metaphysical to epistemological language.

190 Jackson, "Armchair Metaphysics," 32.

191 Ibid, 26.

192 Jackson, *Metaphysics to Ethics*, 7.

- 
- 193 Jackson and Chalmers, "Conceptual Analysis and Reductive Explanation," 316.
- 194 Jackson, "Armchair Metaphysics," 26.
- 195 Jackson on completed science being a whole lot more of the same kind of stuff.
- 196 These examples are all drawn from Chalmers, *Conscious Mind*, 6-11. These six pages provide what are probably the most rich phenomenological descriptions to be found anywhere in the analytic literature.
- 197 See Patricia Churchland, *Neurophilosophy*, 368.
- 198 'Objective' in the sense described in the discussion of first-person knowledge and the problem of demarcation in Chapter Three). That is, an objective phenomenology is the study of subjectivity as an objectively existing part of the natural world. Indeed, the project of European phenomenology dating back to Husserl is to pursue an objective understanding of the lived perspective on the world. Husserl believes that an objective phenomenology is only possible if we suspend the natural attitude and set aside our culturally embedded naïve assumptions about the nature of perception. For an account of the similarity between Husserl's phenomenological reduction and Patricia Churchland's eliminativism, see my "Neurobiology and Phenomenology," 46-49 (reprinted as the Appendix to this dissertation).
- 199 Thusfar, I have generally followed Jackson's account of the entailing facts in terms of 'physical' facts. Having critiqued this ambiguous use of 'physical' that is found in his own writings, I will henceforth largely adopt the use of 'microphysical' to describe the entailing facts. As mentioned in Chapter Two, Chalmers thinks that the knowledge argument ought to be recast in terms of Mary's complete pre-release microphysical knowledge. In articles on conceptual analysis co-authored by Jackson and Chalmers, the entailing facts are similarly described as 'microphysical'. I adopt the use of 'microphysical' here in an attempt to set to the side my claim that Jackson's use of 'physical' is question begging. Nevertheless, as I already argued, the entailment thesis is less plausibly a commitment of physicalism when the entailing facts are 'microphysical' as opposed to 'physical'.
- 200 Block and Stalnaker, "Conceptual Analysis and Explanatory Gap," 14.
- 201 Ibid, 14.
- 202 Ibid, 14.
- 203 Ibid, 15.
- 204 Jackson and Chalmers, "Conceptual Analysis and Reductive Explanation," 338.
- 205 By the way, Ankata is a bearded dragon, a species native to the Australian desert. 'Ankata' is the aboriginal name for the species, literally meaning "no drink". Since Jackson is also an Australian native, Ankata seemed an appropriate example.
- 206 Hardcastle, "Why of Consciousness," 11.

## Conclusion: Solving the Placement Problem

In Chapter Four I agreed with Jackson that physicalists who are not eliminativists regarding qualitative facts face the placement problem. That is, a complete physicalist metaphysical picture must provide a physical account of qualitative facts. I rejected Jackson's view that entry by entailment is the only possible solution to the placement problem; yet, I offered no specific alternative solution. My primary goal in the Conclusion and Appendix is to suggest such a solution to the placement problem. In this Conclusion, I offer a speculative metaphysical account of qualitative properties and their place in physicalist ontology. I also explain how the Appendix, a reprint of my recent article, "Neurobiology and Phenomenology: Towards a Three-Tiered Intertheoretic Model", published in the *Journal of Consciousness Studies*, addresses the epistemological side of the placement problem by suggesting that the theories and conclusions of European phenomenology can play a critical role in discovering genuine qualitative facts and placing such facts in a broader scientific worldview.

### I. A Speculative Physicalist Metaphysic of Qualitative Properties

My goal in this section is as simple as it is ambitious: to explain what kind of thing qualitative properties are, and to justify their inclusion in physicalist ontology. Of course, the critiques I have made of Jackson might be right even if my positive view is wrong. Nevertheless, a speculative account of qualitative properties is the appropriate conclusion to the dissertation for two reasons. First, the precise nature of qualitative



properties is the deepest unexplored issue in the body of the dissertation. Second, my view of qualitative properties is largely informed by my reflections on Jackson's work. Obviously, what follows is more of a snapshot of a process than the presentation of firm conclusions. I have no delusions that my attempts to do metaphysics measure up to the efforts of those who I have thus far dared only critique.

After reiterating the narrow sense in which I believe that physicalism claims to be complete, I will defend the view that each qualitative properties emerges from the interaction between the microphysical properties of the experiencing subject and the totality of the subject's spatial, temporal, and cultural environment.

### *Physicalism and Completeness*

As mentioned in Chapters Two and Four, I believe that Jackson's formulation of the minimal supervenience thesis excellently captures the sense in which physicalists must claim that there are only physical facts. By way of reminder, Jackson's minimal supervenience thesis states that a minimal physical duplicate of our world is a duplicate of our world in every way. In other words, according to the minimal supervenience thesis, a world with the same physical entities, properties, and relations as our world (and with nothing extra added) is an exact replica of our world. I agree with Jackson that, if there are entities, properties, or relations that exist in our world, but fail to exist in a minimal physical duplicate of our world, then physicalism is false in that it provides an incomplete account of the nature of our world. As I also made clear in Chapters Two and Four, I disagree with pre-1998 Jackson regarding whether the qualitative facts about our world are also true in a minimal physical duplicate world, and I disagree with post-1998

Jackson regarding whether there are genuine qualitative facts about our world. My view is that our world contains genuine qualitative properties, and that such properties would also exist in a minimal physical duplicate of our world. In order to justify this claim, it is necessary for me to explain my view that the qualitative properties of our world are best understood as finely grained properties that emerge from the interactions of a wide range of lower-level properties encompassing both interior (primarily neurological) properties and exterior (including cultural) properties.

### *Holism and Supervenience*

Contrary to conventional wisdom in analytic philosophy of mind, I reject the kind of multiple realizability in which higher-level properties are realized (or realizable) by multiple physical properties. Any particular mental event is based on a vastly complicated arrangement of states in many (if not all) parts of the brain. The rich texture of qualitative properties arises from the uniqueness of each human brain, the uniqueness of each person's history, and the uniqueness of each person's collection of social roles. As a result of the richly textured base from which qualitative experiences arise, it seems to me that mental events can only be realized in the way in which they are, in fact, realized. Indeed, the empirical evidence regarding the connection between qualitative experiences and neural states implies an even closer relationship than supervenience. fMRI and PET scan evidence suggests that not only can there be no change in the qualitative experiences without some corresponding neural change, there also can not be a change in certain neural states without a corresponding change in qualitative experience. Not only do qualitative states seem to be ontologically necessitated by

physical states but the particular physical states seem to be ontologically required for certain qualitative states. There appears to be a strict one to one relationship between qualitative facts and neurological facts.

European phenomenologists have long recognized that it is the whole body, historically and culturally located, that engages in perception. Though the various sensory modalities of any perceptual act can be peeled off and examined individually, the lived perceptual experience is experienced as a coherent whole whose character cannot be fully determined by any of the particular sensory modalities nor, and less so, separated from the historical and cultural horizon in which the experience takes place. The blind person whose hearing has acquired unusual acuity hears both more and less in the orchestra than does the sighted person who can observe the interaction of conductor and musicians. A Jewish refugee and a Christian evangelical look at the same cross and experience different qualitative states. It is not enough for those who reject holism to object that they have different feelings about the same visual experience. There is no raw sensory experience that precedes emotive and cultural interpretation of the experience. Visual sensation always takes place within, and can not be separated from, the interpretation of the visual experience's meaning. Perceptual acts are inherently interpretive. No two living organisms have precisely the same spatial and temporal locations or histories; no two organisms are ever in precisely the same physical state; no two conscious experiences are ever precisely the same.

Naturally, such phenomenologically inspired intuitions also inform my view of artificial intelligences such as conscious computers, silicone Martians, and a rightly connected nation of Chinese minds. I accept what has sometimes been derisively called

biological chauvinism. Artificial intelligence is not possible if, by 'intelligence', we mean human-like intelligence. In order to create human-like mental events, it would be necessary to replicate a human brain. Silicone simply will not have the right kind of relationship to hormones, electrical discharges, genetic influences, inputs from organs etc. in order to create experience that human beings would recognize as their own. I would not go as far as Searle and assert that a computer can have no inner life. I am asserting the narrower claim that the inner life of a non-human organism could never be the same as a human inner life, and it is similar only to the extent that it has a physical substrate similar to the human nervous system. Even if a silicone organism were to be appropriately embedded in a social and cultural milieu (though I have no idea how that might happen), subjective and qualitative experiences that would be recognizable by a human being could only emerge if the interior of the organism were properly arranged so as to have a human experience of that milieu. Because the actual qualitative experiences of human beings emerge from such a rich texture of both internal and external interactions, adequate simulation of the necessary interactions would amount to duplication.

### *Emergence and Supervenience*

The term 'emergent' applies to the broad category of properties that arise when the whole transcends the parts. 'Emergent' is a good word for capturing the relationship between qualitative properties and microphysical properties; that is, qualitative properties emerge (with ontological necessity) from their subvenient base. Consider something as simple as salt (NaCl). The taste properties of salt are not something that exists in addition

to the sodium and the chloride. There could be no such thing as zombie salt that is made out of sodium and chloride but doesn't taste like anything. Zombie salt is not possible despite the fact that there is no apparent necessary relationship between the microphysical properties of sodium and chloride on the one hand and the qualitative experience of tasting salt on the other hand. Economic properties also emerge from a microphysical base; there is no possible counterfactual world that is a complete physical replica of our world yet has a different unemployment rate. Social facts (such as the civil rights movement of the 1960s) are also fully determined by, but not reducible to, the physical properties and relations of our world. I would argue that facts about ethics, the history of evolution, and virtually everything other than particle physics are emergent facts. They are ontologically determined by the facts of microphysics but neither explainable by, nor predictable on the basis of, microphysics. The general characteristic of emergent facts is that the whole has properties that are above and beyond the properties of the individual parts. Qualitative facts seem different than other emergent facts not because they are somehow metaphysically different, but because of the peculiar reflexive nature of qualitative propositions. That is, qualitative facts are propositions expressing the end result of trying to know what it is like to know what it is like.

### *Qualitative Properties and Granularity*

Thomas Polger argues that claims of the multiple realizability of mental states often rest on an equivocation of granularity. For instance, he suggests that whereas Fodor has course grained mental states in mind ('pain' is understood by Fodor generally enough to be experienced by both humans and mollusks), Fodor's argument that pain states are

multiply realized relies on a fine grain understanding of brain states (fine enough that they can not be shared across species). Perhaps, Polger argues, a similarly course grained understanding of a brain state might be identical to the mental state that Fodor has in mind. Though for different reasons, I agree with Polger that issues of granularity are central to explaining why mental states are not multiply realized. As already mentioned, I hold to a hold to a holistic view of qualitative properties in order to consider the qualitative experience in its full context; I have fine grain qualitative states in mind. Similarly, the subvenient base from which qualitative properties emerge also needs to be understood on a fine grain. After all, not only are cortical states relevant to qualitative experience, but qualitative experience is shaped simultaneously by virtually every region of the brain, every perceptual tool available, and the entire horizon of one's experience. Put differently, I agree with Andy Clark's characterization of the extended mind. Through the brain, experience reaches out into the so-called external world and the world reaches back in, forming the contours of inner life. An accurate account of the mind must include the extended world in which the mind finds itself. Clark's intuition about the extended mind, itself supported by European phenomenology, breaks down the distinction between the experience and the thing experienced, between subject and object. The world itself is part of the perceptual apparatus by which the world is perceived.

### *Causation and Reduction*

My version of emergent physicalism is consistent with mental causation, but not without some paradox. I believe that qualitative states are causally efficacious through their subvenient base. That is, whereas the qualitative experiences that I had the last time

I was at the symphony is causing my current behavior of buying tickets for next weekend, it is because of the physical basis of my previous qualitative experience that mental causation does not violate the causal closure of the physical world. Substantial paradoxes persist, no doubt. For instance, since I hold that the context of my experience is part of the mechanism by which qualitative properties cause some of my behaviors, the physical composition of the walls in the concert hall, for instance, turns out to be a (small) part of the cause of my current ticket purchase. This perhaps will require me to abandon the causal closure of physics, but not of the physical. So be it; physics is a natural science – an epistemological enterprise and not a metaphysical position. There is nothing spooky about claiming that no theory of physics is going to be able to completely capture the causal intricacies of the world. Causal closure, when rightly understood as a commitment of physicalism, rests on a larger picture of the super- and sub-venient bases, not a claim that is purely about the subvenient level. Physicalists must claim that all causal forces are bound up in a supervenience relationship with the properties described by microphysics, not that the propositions of microphysics can characterize all causal processes. Properties that emerge from microphysics, such as qualitative properties, are bound up in such a supervenience relationship and can therefore be recognized as causally efficacious by physicalists.

The sense in which qualitative properties should, and should not, be considered as standing in a reductive relationship with microphysics is both a final important issue for consideration, and a useful summary of my speculative position. On the one hand, I reject ontological reduction. Qualitative states can not accurately be said to be nothing more than certain complex neurobiological states; qualitative properties must be taken

seriously in their own right. On the other hand, I endorse reductive epistemological strategies as one tool among many for seeking an explanation of consciousness. Explanations of higher-order phenomenon are best pursued by grounding them in lower-order phenomenon. Endorsing epistemological reduction, while rejecting metaphysical reduction, might seem more than counter-intuitive. Why seek a reductive understanding of something that you don't even think reduces? But the paradox here is merely apparent, resting on an equivocation regarding 'reduction'. Rejecting ontological simplification in no way diminishes the value of intertheoretic connections between various explanatory levels. Indeed, it is precisely the establishment of intertheoretic connections between phenomenology and the traditional natural sciences that buttresses my claim that qualitative properties supervene on microphysical properties.

## II. A Speculative Physicalist Epistemology of Qualitative Properties

The purpose of this section is to explain how the Appendix relates to the rest of the dissertation. In the preceding section I provided a speculative metaphysical account of qualitative properties. The Appendix is a reprint of my "Neurobiology and Phenomenology: Towards a three-Tiered Intertheoretic Model of Explanation," recently published in the *Journal of Consciousness Studies*. As the stated thesis of that paper is quite distinct from the content of this dissertation, it is necessary for me to explain how it fits in with the rest of the dissertation. In short, I offer the reprint as a speculative strategy for coming to know qualitative facts and for understanding the relationship between phenomenology and the traditional natural sciences.



### *The Need for an Epistemological Account*

Two questions might rightly be asked of the speculative metaphysical account outlined above. First, one might ask if there is anything to ground my intuition that the qualitative facts in our world would also obtain in a minimal physical duplicate world. Second, one might ask whether my description of physicalism is so broad as to make the proposition that ‘we live in an entirely physical world’ vacuous; put differently, given that I have defined ‘physical’ broadly enough to include qualitative properties among the physical properties, one might suggest that I have defined ‘physical’ too broadly to exclude anything at all. In the sense that my speculative metaphysical account of qualitative properties does not provide the tools to answer these questions, my metaphysical account is unable to stand on its own.

I will offer a broadly epistemological response to both of these related questions. As I see it, the best evidence that qualitative properties supervene on, and emerge from, lower-level properties is that the results of systematic investigation of qualitative properties stand in an intertheoretic explanatory relationship with the rest of the natural sciences. To the extent that findings about qualitative properties and findings about neurobiological make sense in light of one other, my intuition that qualitative properties have a place in physicalist ontology is strengthened.

### *Summary of the Appendix*

The stated thesis of the paper reprinted as the Appendix is that collaboration between analytic and continental philosophies of mind is possible, useful, and overdue. After surveying the current state of the emerging conversation between the two traditions,

I suggest a theoretical model of explanation in the study of the mind that integrates the work of phenomenologist Maurice Merleau-Ponty and analytic philosopher of science Patricia Churchland. I point out that Merleau-Ponty's account of the relationship between phenomenology and psychology is surprisingly similar to Churchland's account of the relationship between psychology and neuroscience. On both accounts, the lower and higher levels of discourse serve to inspire and constrain one another. Both theorists could be open to the suggestion that phenomenology, psychology, and neuroscience can be understood in a three-tiered relationship in which theories generated from any of the three tiers might play a role in the development of theories at either of the other two levels. Such a role might be positive or negative. For a positive example, neurobiological discoveries might offer explanatory patterns useful to phenomenology; for a negative example, phenomenological observations might demonstrate that a certain line of neurobiological research is misguided. As a specific example to ground my suggestion, I examine neural plasticity. Neuroscience demonstrates that localization of cortical functioning is determined by sensory input; psychology demonstrates that cognitive functioning can be rewired following localized resection; phenomenology demonstrates that the subject constructs itself through its interaction with the world. Citing existing research in neuroscience, clinical psychology, and phenomenology, I try to demonstrate that my suggestion for a three-tiered intertheoretic model is both a useful way to understand what has already been discovered across all three levels of investigation and is a useful tool for future research in neuroscience, psychology, and phenomenology.

The terms 'qualitative facts,' 'physicalism,' and 'placement problem' are generally not used in the paper reprinted as the Appendix. Nevertheless, my suggestion of

a three-tiered intertheoretic explanatory model addresses the placement problem in that the suggestion places qualitative facts within the explanatory framework of the natural sciences – an explanatory framework that is universally, and without controversy, believed to be consistent with physicalist ontology.

### *The Appendix and the Placement Problem*

By way of reminder, the placement problem is faced by any metaphysical viewpoint that seeks to understand the world in terms of some limited number of fundamental ingredients. The version of the placement problem faced by physicalists is that there are a whole range of putative properties of our world that are not explicitly among the fundamental physical ingredients of our world (psychological properties, theological properties, even biological properties). Physicalists must either hold that such properties are merely putative (that is, take an eliminativist stance regarding the properties in question) or place such properties in the story generated by the fundamental physical properties (that is, show that the properties in question are made true by the fundamental physical properties).

In addition to encouraging more collaboration between analytic and continental philosophies of mind, the paper reprinted as the Appendix argues that systematic investigation of qualitative facts is not mere wishful thinking about some future enterprise. It is the ongoing project of European phenomenology, with a coherent methodology and a broadly accepted theoretical framework that dates back at least to Edmund Husserl. Of course, phenomenology and neuroscience have each developed with researchers who are largely unaware of (and often overtly unconcerned with) what is

happening in the other field. That the two fields have independently developed interconnecting patterns of explanation provides, it seems to me, compelling evidence that qualitative properties are not features of our world that are entirely independent of the properties described by neuroscience. At the very least, it would be an amazing coincidence if qualitative and neuroscientific properties succumbed to similar patterns of explanation without there being a deep ontological connection between the two sets of properties. Admittedly, such a connection may or may not be a reflection of the particular speculative picture of qualitative properties that I outlined above, but the existing explanatory connections indicate that some such picture almost certainly underlies the actual structure of our world. Put differently, while the empirical evidence does not demonstrate that I have provided an accurate account of how the qualitative properties are made true by the fundamental physical properties, the empirical evidence does suggest that the fundamental physical properties make the qualitative properties true *in some way or another*.

Physicalists who are not eliminativists about qualitative properties ideally ought to solve the placement problem by providing a description of the nature of qualitative properties themselves. Nevertheless, it seems to me that physicalists who wish to take qualitative properties seriously will have sufficiently addressed the placement problem if they are able to show that explanatory patterns and epistemological strategies suitable for the chemical and biological sciences (such as neuroscience) are also suitable for systematic investigation of qualitative properties. The appeal of physicalism is that physicalism claims to offer a unified naturalistic picture of the world that we live in. By offering a unified picture of the world, physicalism avoids the intractable problems

associated with substance dualism. By offering a purely naturalistic picture of the world, physicalism offers an optimistic view of the potential extent of human understanding that dramatically contrasts with the fundamental mysteries proffered by more traditional metaphysical viewpoints. Physicalists who repudiate eliminativism regarding qualitative properties ought to be credited with making good on their promises of a unified and naturalistic explanation if they can demonstrate that the naturalistic methodology that has so effectively offered insight into various phenomena along levels of explanation ranging from microphysics to (at least some) psychology can be fruitfully and smoothly extended to phenomenology. Put much more simply, if the pursuit of qualitative facts has a firm place in the epistemology of the natural sciences, concerns about whether qualitative properties can be placed in a physicalist ontology are greatly mitigated.

Which leaves me the task of directly addressing the two related questions that I acknowledge might rightly be addressed to the speculative metaphysical account of qualitative properties are unique emergent properties that I offered in the previous section. As to the first question, the grounding for my intuition that the qualitative facts in our world would also obtain in a minimal physical duplicate world is, quite simply, the empirical evidence. Just as I believe that biological properties supervene on lower-level properties because the explanatory patterns useful in biology interconnect with those that are useful in chemistry and physics, so too I believe that qualitative properties supervene on the properties described by the traditional physical sciences for precisely the same reason. As to the second question, my description of what counts as a physical property is narrow enough to exclude any putative properties of our world that can not be brought into explanatory cohesion with the natural sciences. For an obvious example, most people

believe that our world was designed and created by a supernatural God. As any explanation of such a being would necessarily stand outside any possible theory of physics, chemistry, or biology, I must respectfully disagree with most people about the existence of such a being. If there are theological properties of our world, they were better described by Spinoza than by St. Paul. My description of what counts as a physical property is also narrow enough to raise deep suspicion about putative properties of our world that seem highly unlikely to be brought into explanatory cohesion with the natural sciences. For example, those who wish to persuade me that animals (such as human beings) are capable of ESP will make no progress until they can provide some evidence of a chemical or biological mechanism making such powers possible.

The overall spirit of physicalism might be characterized as the hard-nosed refusal to acknowledge the existence of anything spooky. Though dualist accounts of mental life are unavoidably spooky, the proper physicalist response is not to undermine the reality of qualitative experience. Instead, physicalists ought to seek an account of qualitative states, both metaphysically and epistemologically, that focuses as clearly as possible on the objective existence of qualitative experience as a natural phenomenon in the world. We need to seek, in the universally valid structures of our own subjective experiences, the objective facts about subjective experience. As Edmund Husserl proclaimed, “to the things themselves.”

## Appendix

Boyle, Noel. "Neurobiology and Phenomenology: Towards a Three-Tiered Intertheoretic Model of Explanation." *Journal of Consciousness Studies* 15, no. 3 (2008): 34-58.

Neurobiology and Phenomenology:  
Towards a Three-Tiered Intertheoretic Model of Explanation

Analytic and continental philosophies of mind are too long divided. In both traditions there is extensive discussion of consciousness, the mind-body problem, intentionality, subjectivity, perception (especially visual) and so on. Between these two discussions there are substantive disagreements, overlapping points of insight, meaningful differences in emphasis, and points of comparison which seems to offer nothing but confusion. In other words, there are the ideal circumstances for doing philosophy. Yet, there has been little discourse.

This paper invites expanding discourse between these two philosophical traditions. The first part briefly describes the existing literature which works across the analytic-phenomenology divide, situating my work within it as a focus on analytic physicalism and phenomenal explanation. In the longer second part, I sketch a model for explanation embedded simultaneously in both traditions. Hopefully, a theoretical framework emerges that the unlikely combination of Maurice Merleau-Ponty and Patricia Churchland *could* accept. In the third part, I apply the three-tiered model to a discussion of neural plasticity and suggest that the model both reflects existing research across three levels of analysis and can be a fruitful way to approach future research.

My suggestion for a three-tiered model is quite tentative. Much less tentative is my claim that constructive dialogue between phenomenological and physicalist study of consciousness is long-overdue, illuminating, and practical.



## **I. The Current Situation**

Since Thomas Nagel's landmark essay, 'What it is like to be a bat?' (1976), analysis of subjective experience has become central to analytic philosophy of mind. Emerging from a variety of related sources (Nagel, 1976; Kripke, 1980; Jackson, 1982; Chalmers, 1996), much debate in analytic study of consciousness in the last three decades has focused on the first-person point of view; subjectivity must be taken seriously if we are to account for 'qualia'. Whereas phenomenology is focused on the lived-world (*Lebenswelt*), analytic philosophy of mind has recently turned its attention to 'what it is like to be' in a certain mental state. Though the terminology is distinct, the same phenomenon seems to be under investigation. In both cases, the emphasis is on subjective experience. Near the end of Nagel's essay, he speculatively proposes that objective descriptions of subjective experience can be sought. It might be possible, he suggests, to give objective description of the content and structure of lived experiences. We would need a new conceptual framework, new methods of investigation and 'we would eventually reach a brick wall' but he clearly indicates it would be fruitful to see how far we could go. Though not surprising, it is unfortunate that such a prominent philosopher could write this 'speculative proposal' and not recognize it as the dominant project of the last century of European philosophy. Nagel's essay thus highlighted both the potential value and nearly complete absence of analytic-phenomenological co-operation.

Many attempts to work across the analytic-phenomenological divide in the philosophy of mind are not exactly co-operative. Too often, one tradition is simply criticized from the standpoint of the other. While such critiques can open new challenges to even hidden assumptions, they generally do so at the cost of reinforcing, and not

overcoming, the isolation between analytic and phenomenological philosophy (see Dennett, 1991, p. 44 and Carr, 1998 for a particularly unfriendly exchange). In better cases, one tradition offers interpretive or methodological advice to the other (see Wilder, 1997; Kelly, 2001). While this is also very helpful, and perhaps a necessary first step, it is something short of bringing both traditions to bear in understanding the mind. Ideally, there should be an approach to the philosophy of mind *unifying* the two traditions in one coherent discourse.

In recent years, a vast literature seeking unification of phenomenology and analytic philosophies of mind has emerged. Francisco Varela (1996) coined the term ‘neurophenomenology’ and argued that phenomenology offers useful methodological tools for addressing, as David Chalmers (1995) called it, the ‘hard problem’ of consciousness: the subjective, qualitative, ‘what it is like’ aspect of experience. Agreeing with much of Chalmers’ assessment of the failure of reductive approaches to consciousness, Varela suggests that the something fundamentally different needed to explain consciousness is phenomenological description. Contrary to Nagel’s assessment that this would eventually hit a brick wall, or Dennett’s suggestion that phenomenology has no unified method and thus no worthwhile results, Varela points out that European phenomenology has an uninterrupted methodological history dating back to Husserl. He also effectively debunks common myths within analytic philosophy regarding the rigor and ‘introspectionism’ of phenomenological analysis. (p. 338). While it is certainly not a settled issue that first-person accounts can offer scientifically credible evidence, he persuasively argues that those who take seriously the first person perspective ‘must inescapably attain a level of mastery’ in the phenomenological method of description (p.

347). If one understands that the first person account is indispensable, one must acquire the descriptive skills necessary to mine it. Though this still largely amounts to offering phenomenological advice to analytic philosophy, Varela's work is at least rooted in both traditions in the sense that a problem arising within analytic philosophy is offered a phenomenological solution; Varela does not use phenomenology to dictate a problem set to analytic philosophy.<sup>1</sup>

Locating the role of phenomenology within analytic approaches to the mind, Varela summarizes analytic philosophy of mind with a 'four way sketch', indicating that those analytic philosophers who hold that a first-person account is central (Chalmers and Searle are the most discussed in the article) are committed to something like a phenomenological methodology (333).<sup>2</sup> Varela's sketch leaves the impression that physicalist approaches, especially reductive ones, are beyond phenomenological help. On his map of analytic philosophers, he places Churchland as far from the arch of phenomenological relevance as is possible. Varela describes Churchland's eliminative materialism as an attempt to 'eliminate the hard problem by eliminating the pole of experience in favor of some form of neurobiological account' (333). I argue in the next section that this is an (all too common) oversimplification and misreading of Churchland's<sup>3</sup> position. More generally, I do not agree with the implication that

---

1. See Bickle & Ellis (2005) for a competing account of the claim that phenomenological methodology is useful for addressing the 'hard' problem.

2. Nagel is a notable omission from those considered by Varela to take seriously the first person perspective. Instead, Varela locates him near 'mysterianism' probably because Nagel suggests that because we can not have a first person perspective except in our own case, we simply can not know what it is like to be, for instance, a bat.

3. There are, of course, two Churchlands. On my reading, they are entirely in agreement and the apparent minor differences between them reflect both a difference of interest and a higher degree

phenomenology has no meaningful point of dialogue with physicalist and neurobiological approaches to consciousness. Therefore, one primary goal of this paper is to argue that interaction between phenomenologists and physicalist oriented analytic philosophers is both possible and valuable. Even reductive approaches, properly understood, are within the arch of phenomenological relevance.

By emphasizing neurobiological approaches to consciousness, this paper adds to a growing and very recent body of literature recognizing the extent to which emerging neuroscience supports, helps to explain, and illuminates core phenomenological insights (see Stawarska, 2003; Overgaard, 2004; Gallagher, 2005; Bickle & Ellis 2005). However, existing work tends to focus on specific points of comparison between neurobiological findings and phenomenology instead of attempting to assess the ‘big picture’. Further, there is little engagement with the analytic philosophy of science through which theorists such as Churchland evaluate and interpret neurobiological findings. Therefore, questions about the structure of a unified explanation as such have not been thoroughly explored. As a result, little has emerged regarding a unified research agenda which can accommodate physicalist and phenomenological intuitions. Therefore, the other goal of this paper is to suggest that resources exist to articulate a model of explanation for the mind which contains respectable places for phenomenology, psychology and neurobiology.

## **II. A Three-Tiered Model?**

---

of carefulness on Patricia’s part. For both of these reasons, I rely almost entirely on her work in this paper. So, ‘Churchland’ refers to Patricia Churchland, unless otherwise noted.

In broad strokes, I propose a three-tiered intertheoretic model of explanation in which there is explanatory coherence across three levels of investigation: phenomenology, psychology and neuroscience. Therefore, phenomenology and neuroscience are brought under the same explanatory framework without either collapsing into the other. The general idea is that the most important aspects of Churchland's and Merleau-Ponty's philosophies can be 'fit together'. I am seeking a model based on cooperation, not competition. Thus, I hope to I offer a model that could (not necessarily would) be accepted by Merleau-Ponty and Churchland simultaneously, without either of them renouncing their most basic philosophical positions.

*A. The immediate metaphysical problem*

Metaphysically, Churchland and Merleau-Ponty might be described (broadly representative of their respective traditions) as two different approaches to rejecting Cartesian dualism. Largely motivated by the refractory nature of the mind-body problem, each embraces a form of monism, overturning the historical distinction between mind and body. For both, 'mind' is necessarily embodied. Here metaphysical agreement apparently ends and a great rift seems to open up.

Churchland is a hard-nosed materialist; the mind-brain is a material thing. There is no ghost in the machine; there is no immaterial soul. She is the intellectual heir of the identity thesis. The 'mind-brain' is a natural phenomenon, available for scientific examination. Wild speculation about disembodied minds belongs in the metaphysical dustbin, along with phlogiston and vital spirits. Merleau-Ponty, however, is not a materialist. He emphasized that the mind does not inhabit the body, but that the body is

the point of view on the world and consciousness is necessarily embodied. There is an identity thesis here, but it is not identity of brain states and mental states but of the subject and the subject's embodiment. Replacing the dualist language of 'mind and body', Merleau-Ponty develops the notion of 'the flesh of the world', the reversible structure of subject and object through which consciousness is instantiated. But, 'the flesh we are speaking of is not matter' (1968, 146); it is active, and interactive, the seat of consciousness - not mere matter. He rejects an idealized subject or a purely objectified matter, and takes an interest in the lived world. That which is 'given to experience' is the phenomenon itself. Empty abstractions like disembodied minds belong in the metaphysical dustbin, along with 'things as they actually are, independent of our experience'.

While there is a real difference regarding what is to be studied, the methodological and practical implications of the difference are limited. It is a disagreement that can exist *within* the model I sketch. In order to see this, two observations are crucial. First, Merleau-Ponty does not reject the legitimacy of scientific investigations; the sciences have an autonomous role which stands in a certain relationship with phenomenological investigations. To say, 'the flesh we are speaking of is not matter' does not mean flesh is some kind of non-matter, or that the material sciences are irrelevant but only that the cold, dead descriptions of 'atoms in the void' are woefully inadequate. Second, for Churchland eliminative materialism is an empirical prediction regarding folk psychological theories, and not an ontological commitment to the nonexistence of higher levels of reality. As the next section explains in detail, Churchland does not suggest that only the lowest levels of explanation are legitimate.

There is no metaphysical reason that she has to rule out phenomenological investigation as part of the explanatory picture.

Though substantial, their metaphysical disagreements do not rule out collaboration and explanatory unity between neuroscience, psychology, and phenomenology. One could hold to either a materialist or a phenomenological ontology *and* seek explanatory unity. So: let the materialism - idealism debates rage on. Either ontology can motivate a broad statement of epistemic strategy: *examine the unified whole that is present and available for systematic examination*. And, for present purposes, that is enough.

#### *B. Inter-theoretic relations*

Churchland extensively describes the intertheoretic relationship between psychology and neuroscience. Merleau-Ponty extensively describes the intertheoretic relationship between phenomenology and psychology. These two accounts are similar enough (in the relevant ways) for them to be brought together into a unitary explanatory framework which has Merleau-Ponty's phenomenology at the highest level and Churchland's neuroscience at the lowest level, with psychology sitting stably in the middle. Recognizing that this conclusion is perhaps surprising, I begin my argument for it with a careful reconstruction of Churchland's model of intertheoretic reduction. Contrary to popular oversimplifications, Churchland does not eliminate psychology and reduce everything to neuroscience. Churchland suggests that psychology and neuroscience will co-evolve and *some future theory of psychology* will stand in a reductive relationship with *some future theory of neuroscience*. In 'Do We Propose to Eliminate

Consciousness', Churchland summarizes the conventional critique of her view that it can ultimately say little about consciousness because it tells only of neural discharges, and calls it 'straw through and through'. She suggests that the proper response to the objection is 'to embrace it' and recognize that neuroscience can only be a part of a coherent explanation. In other words, Churchland's eliminativism extends only to *folk psychology* (common sense, prescientific, psychological theory): folk psychology will be eliminated<sup>4</sup> and, in its place, a *scientific psychology* will emerge in reductive co-evolution with neuroscience. In this section, I emphasize her positive view of the relationship between scientific psychology and neuroscience. I then turn to Merleau-Ponty's view of the relationship between psychology and phenomenology.

#### --*Neurophilosophy*

In *Neurophilosophy*, Churchland outlines a model of intertheoretic reduction which builds from, but substantially modifies, the classic account by Ernest Nagel (1961).<sup>5</sup> Nagel sees intertheoretic reduction as a particular form of deductive-nomological (D-N) explanation, in which a particular phenomenon is said to be explained when its occurrence can be deduced from a combination of relevant general scientific laws and specific empirical observations (see Hempel, 1965). On Nagel's account, one

---

4. The folk psychological concept of 'memory' has been largely fractured by studies of patients like HM and Korsakov's sufferers, demonstrating that 'memory' does not denote a single capacity of the brain. Split brain studies primarily by Sperry have similarly shown that 'self' is not a unified phenomenon. Studies of blindsight have similarly fractured the folk psychological concept of 'seeing'. She points to these and similar insights from the emerging scientific psychology.

5. See Schaffner 1993 and McCauley 1996 for accounts of Churchland's adaptation of Nagel's account.



theory is said to reduce to another when the reduced theory follows deductively from the reducing theory with the aid of 'bridge laws' translating the conceptual language of the reduced theory into the conceptual language of the reducing theory. On this model, reduction is the kind of D-N explanation in which the occurrence of the explained phenomenon follows deductively from general laws and specific empirical facts that are at a lower level of physical reality.

Churchland's most important improvement on this model is to recognize that both the reduced and reducing theories generally need to undergo some degree of revision before 'reductive consummation' (1986, p. 283). Instead of doing all of the scientific investigation first and sorting out the reductive relationships later, the more practical approach (as well as the approach of practicing scientists) is for theories at neighboring levels of physical reality to inform and constrain each other as they progress. No doubt, there is not currently a reductive explanation of mental phenomena, and none is forthcoming from our current understanding of either the mind or the brain. Thus, the candidate for reduction to neurobiological theory is not the theory of the mind as we currently understand it. Instead, it is 'the integrated body of generalizations describing the high-level states and processes and their causal interconnections that underlie behavior...the domain of scientific psychology...some *future* theory' (1986, p. 295).

Thus, by advocating mind-body reduction, Churchland is not suggesting that the mental is an illusion and there is 'only neuroscience'. On the contrary, the main thrust of her book is to advocate a co-evolutionary research program in which psychology and neuroscience work together towards theoretical unification in which mental states are explained in terms of brain states. What comes of our current set of psychological

concepts (or neuroscientific concepts, for that matter) is an empirical matter, to be determined in the process of the movement towards reductive consummation.

As a result, there is a continuum of possible outcomes for our current theories of psychology and neuroscience, ranging from a straightforward reduction approximating the intertheoretic relationship envisioned by Nagel, all the way to a complete elimination of the current theory (along with its conceptual framework) in favor of one that is capable of reductive consummation.<sup>6</sup> Though it is important to remember that this is a continuum, it is useful to distinguish between three different points on the continuum to highlight the range of possible outcomes for current theories in psychology and neuroscience. First, there might be a 'smooth reduction', that is, reduction as Nagel conceives it. Moving partway across the continuum, there might be a 'bumpy reduction' in which substantial but not wholesale revision of the reduced theory takes place on the way to reductive explanation. In other words, what gets reduced is not the higher level theory per se, but an updated and improved version of the theory. Completing the move across the continuum, it might turn out that the current theory is so flawed that it simply needs to be eliminated and replaced in order to achieve a reduction. For instance, as late as the 19th century, epileptic seizures were generally explained by a theory of demonic possession. It is only once this demonic possession theory was completely abandoned that seizures

---

6. In fact, it might turn out that the reduction of psychology to neuroscience is not possible, either for practical reasons such as excessive complexity of the intertheoretic relationship, or even in principle. Churchland emphatically does not deny that it is possible there is no reduction between psychology and neuroscience to be had. The point is that denial of such an intertheoretic relationship is an empirical claim for which there is little evidence, given the current infancy of both sciences involved. Their concern is 'only to rebut the counsel of impossibility' (Churchland and Churchland 1990).

could be explained reductively, in terms of abnormal patterns of neuro-electrical discharges.<sup>7</sup> This ‘elimination’ can, broadly speaking, take two forms. The theory can be eliminated either because it points to an illusory level of reality (in which case the original theory was not a theory of anything real) or the existing theory can be eliminated and the same level of physical reality comes to be explained under a new theory (in which case the original theory was a bad theory of something real).<sup>8</sup> ‘Eliminativist’, therefore, refers only to a specific part of Churchland’s empirical predictions and sidesteps her overall structure of her position. Indeed, to reduce Churchland to ‘eliminativism’ is to eliminate the best parts.

Working within a similar model of inter-theoretic reduction, a recent article by Dan Steel asks ‘Can a reductionist be a pluralist?’ (2004). Answering in the affirmative, Steel describes how two sciences which are in an intertheoretic reductive relationship must maintain a degree of autonomy. For instance, if psychology is to be in a reductive relationship with neuroscience then there must be properly psychological investigations in order to develop a theory of psychology which can then connect with a similarly developed theory of neuroscience. As Kenneth Schaffner points out in an oft-cited article

---

7. This last possibility highlights P.S. Churchland’s other, and related, improvement on Nagel’s model: the rejection of bridge laws. In those reductions on the retentive or smooth side of the spectrum, bridge laws will be established. In those reduction on the eliminativist end of the spectrum, it is absurd to suggest that bridge laws are necessary. Consider the epilepsy example. It would be absurd to say that a reductive explanation of a seizure must include laws for translating ‘demonic possession’ into some particular brain state. Yet under Nagel’s model, this is exactly what late 19th century neurological research should have been geared towards.

8. An example of the former is the reduction of the realm of heavenly bodies to Newtonian mechanics (celestial bodies do not represent a distinct level of reality) (see Churchland and Churchland 1996, p.223); an example of the latter would be the epilepsy example (seizures certainly are real).

(1993), reductive explanations point toward a unity of scientific understanding in which theories and sub-theories at multiple independent levels of investigation cohere in an explanatory whole, in which explanations of any given phenomenon are sought out in the terminology of the next lowest level of investigation. Thus inter-theoretic reduction is quite different than Steven Weinberg's 'dream of a final theory' in which science develops to the point where there is nothing but the lowest level of investigation, where other sciences dissolve in the face of the explanatory power of some idealized physics. From a practical point of view, the essence of inter-theoretic reduction is to let a thousand research projects bloom and let each be aware of what the others are doing. It is an explanatory strategy, a matter of epistemology. Unlike Weinberg's dream of microphysical reduction, it is not a matter of metaphysics.

Todd Grantham (2004) has suggested that this unity of science is better characterized as 'interconnection' than as 'reduction'. Given that there are too many relevant senses of the word 'reduction' to keep track of, I will follow his usage. When I refer to an intertheoretic reduction in which higher and lower levels of explanation maintain their autonomy, I will use 'intertheoretic interconnection'. When referring to a metaphysical relationship in which the higher-level phenomenon is 'nothing but' the lower-level explanation, I will use the term 'ontological simplification'. I will leave the phrase 'phenomenological reduction' alone.

*--Phenomenology and the Sciences of Man*

Following Merleau-Ponty's account (itself derived from Husserl), it is reasonable to describe the relationship between phenomenology and psychology as inter-theoretic interconnection.

Husserl originally maintained a priority of philosophy over psychology but, 'as his thought matured, this relation of priority gave way to one of interdependence and reciprocity' (Merleau-Ponty, 1964a, p. 94). Merleau-Ponty even goes so far as to say that 'the conflict between systematic philosophy and the advancing knowledge of science must stop' (1964b, p. 44). Merleau-Ponty focuses on the top-down constraints and insights from phenomenology for a scientific psychology. For instance, he writes,

'When a psychologist speaks of consciousness, the mode of being he attributes to it does not differ radically from things. Consciousness is an object to be studied, and the psychologist sees it among other things as an event in this system of the world. To arrive at a conception which will do justice to the radical originality of consciousness, we need an analysis of a very different type, which will find in our experience the meaning, or the essence, of every possible *psyche*.' (1964a, p. 58)

Phenomenology can provide insight into the nature of consciousness and, on the basis of this insight, empirical psychology can both design and interpret experiments in a way that is more fully grounded in an understanding of the radical nature of consciousness.

Merleau-Ponty also makes some specific suggestions for empirical psychology's handling of images. He critiques a tendency in psychology to regard the image as, 'a little frozen picture in consciousness' (1964a, p. 60) because this approach does not consider

the image within the context of the world to which it relates. On the contrary, Merleau-Ponty writes, 'to perceive oneself as imagining is to set up an operation of my whole consciousness' (1964a, p. 60). Merleau-Ponty then suggests that this recognition might inspire investigation into how the subject 'achieves this incantation of an absent visage in the present data of his perceptions' (1964a, p.60). He speculates that there might be some neural mechanism discovered by which the subject projects a previously acquired structure of the perception by his 'motor-affective attitude'. 'Such an eidetic analysis of the image will make possible experimental approaches which are no longer blind, because they will know something of what they are talking about and will understand the connection of the image with our motor-affective life' (1964a, p. 60). In other words, psychology has much to learn from phenomenology.

A more difficult question is whether phenomenology can recognize bottom-up constraints from scientific psychology. Merleau-Ponty offers no such examples and Husserl's rejection of psychologism gives one great reason to pause. Psychologism is the tendency within psychology to reduce all mental life, including philosophical reflection, to one's psychological state in such a way that the philosopher does not apprehend anything essential about the world but merely reflects psychological biases and determinations. There is also a sociologism, an anthropologism, a historicism etc. Any of these 'isms' of the social sciences make phenomenology irrelevant; if phenomenology just reflects social, psychological, and historical determinations then it does not provide genuine insight into the essence of experience. As Merleau-Ponty put it, 'it is essential that our life should not be reduced exclusively to psychological events and that in and through these events there should be revealed a meaning which is irreducible to these

particularities' (1964a, p. 53). This is a transcendental argument against the ontological simplification of phenomenology to psychology. However, we need to avoid equivocations regarding 'reduction', and the claim that Churchland endorses reduction while Merleau-Ponty rejects it is misleading. The rejection of psychologism is an assertion of the autonomy of phenomenology as a field of scientific investigation, not a refusal to co-operate with psychology. The rejection of psychologism reinforces, not undermines, a model of intertheoretic interconnection.

Thus, Merleau-Ponty's account of the intertheoretic relationship between psychology and phenomenology coheres with Churchland's account of the relationship between psychology and neuroscience. Though they are each critical of psychology (and folk psychology) for reasons that seem to run counter to one another, we might describe them both as suggesting that in order to be a science psychology needs to be part of a larger, and unified, explanation. Merleau-Ponty suggests that psychology is not sufficiently informed and constrained by phenomenology; Churchland suggests that psychology is not sufficiently informed and constrained by neuroscience. Equally, Merleau-Ponty suggests that phenomenology has much to learn from psychology; Churchland suggests that neuroscience has much to learn from psychology (though neither offers much by way of example). And there is nothing in either of their comments about intertheoretic relationships which is counter to the suggested extension of their comments to a third-tier.<sup>9</sup>

---

9. On the contrary, both Churchland and Merleau-Ponty point out that the two-tiered intertheoretic accounts that they present are oversimplified in that there will likely be levels which intervene between the two they focus on.

### *C. Parallel dangers, a mutual solution*

The standard critique of Churchland is that she advocates a bottom-up strategy in which psychology is overrun by neuroscience. As I have argued, this critique is too simple. In order to understand its true force, it must be recast. She explicitly says there must be top-down constraints on neuroscience which will be provided by psychology. Without such constraints, neuroscience will have no guidance. Nevertheless, she is still open to the critique that the explanations which she favors and advocates rely too heavily on neuroscience and do not take top-down constraints from psychology seriously enough. Especially given her predictions that folk psychology will be eliminated, and her claim that scientific psychology is in its infancy, there is a danger that the co-evolution of neuroscience and scientific psychology will be unbalanced, favoring constraints from the bottom-up (neuroscience constraining psychology) over those from the top-down (psychology constraining neuroscience). In other words, any scientific psychology which might constrain neuroscience will be pegged for 'elimination', and psychology will be so overrun with 'constraints' from neuroscience that it will be unable to take up its proper task. More specifically, when psychological investigation yields conclusions which suggest theory modification in neuroscience, there must be a (non-empirical?) decision made between a) elimination of psychology's conclusions and b) further investigation that acknowledges a top-down constraint. Churchland's intense interest in neuroscience (and not psychology) lends credibility to the suggestion that she would invariably choose the former.

There is a similar danger in Merleau-Ponty's description of the relationship between phenomenology and empirical psychology. Any empirical psychology which



might constrain phenomenology will be written off as mere 'psychologism', stuck in the 'natural attitude'. His tendency to describe the project of psychology based entirely on the observations of phenomenology lends credibility to this critique. More specifically, when psychological investigation yields conclusions which seem to suggest a need for reassessment of phenomenological conclusions, there must be a decision made between a) rejecting the psychology's conclusions on the grounds of 'psychologism' or b) further phenomenological investigation which acknowledges a bottom-up constraint. Merleau-Ponty's intense interest in phenomenology (and not psychology) lends credibility to the suggestion that he would invariably choose the former.

A three-tiered intertheoretic model addresses these dangers. Broadly speaking, theories in phenomenology will interconnect with theories in psychology and theories in psychology will, in turn, interconnect with theories in neuroscience.<sup>10</sup> The result, to oversimplify, is that phenomenology protects psychology from being overrun by neuroscience and neuroscience protects psychology from being overrun by phenomenology. In other words, scientific psychology will be constrained from below by neuroscience and constrained from above by rigorous phenomenology. Phenomenology will have a harder time leveling the accusation of psychologism at a psychology that is already in explanatory unification with neuroscience; neuroscience will have a harder time eliminating psychology that is already in explanatory unification with phenomenology. Naturally, the danger with this model is that psychology will be torn apart, with phenomenologists and neuroscientists each having the psychology they favor,

---

10. Of course, there are bound to be more than these three neatly carved levels of investigation, but this oversimplification serves present purposes.

forever talking past one another. Indeed, to a great extent, this seems to describe the current situation. To make the argument that things need not be this way, in the next section I point out that Churchland and Merleau-Ponty *are* sympathetic to each other's central conclusions regarding consciousness.

*D. A surprising synchronicity: eliminativism and phenomenological reduction*

As already described, Churchland suggests that folk psychology will be eliminated, that pre-reflective commonsense psychology will be shown to be an inaccurate description of the mind-brain, impeding rather than elucidating understanding. There are two points to emphasize. First, one can agree with Churchland regarding the proper methodology for seeking an understanding of the mind-brain and disagree with her empirical predictions. One can argue that folk psychology will survive scientific scrutiny. Second, agreeing with the eliminativist thesis does not imply that mental life is an illusion but only that our commonsense explanations of mental life are embedded with misleading illusions.

What is important to Churchland is not whether folk psychology will in fact be eliminated, but that such an elimination be taken as a real possibility. Pre-scientific explanation can, and often does, turn out to be deeply misguided. In investigating the mind-brain (or anything else) we can not assume that our everyday experiences reveal even what there is to be explained. In recognizing the possibility of elimination, we must bracket our naive assumptions.

In phenomenology, where avoiding naive assumptions is more difficult, the task is accomplished by the phenomenological reduction. As Husserl describes it, 'one must

simply break with that supposedly so evident thought stemming from natural thinking, that all that is given is either physical or psychical' (Bernet, 1993, p. 61 – quoting Husserl). Phenomenology's starting point is that the only point of access we have to the world is through subjective experience. Thus, there is always the danger that phenomenology will wallow in subjectivity and never say anything universal. There must be a perspective from which to give objective descriptions of the content of subjective experiences. This is the task of the phenomenological reduction.

The phenomenological reduction is a tool for focusing exclusively on the immediate content of lived experience; it is a bracketing of that which is my own in subjective experience so that the given experience can be approached directly. Thus, one problem with Nagel's speculative proposal is that Nagel still assumes phenomenological knowledge will have to be objective. On the contrary, given that it takes subjective experience as its subject matter, phenomenology must be on guard against the suggestion that there is an objectivity to be had. This is what is meant by phenomenology's rallying cry, 'to the things themselves'. In the phenomenological realm of investigation, the thing itself *is* the thing as it is given to experience. The task of the phenomenological reduction is to focus on the universal structure of subjectivity. 'Husserl metaphorically describes the reduction as a 'method of doffing the empirical – objective robe... with which I again and again drape myself in an habitual apperception that remains unnoticed in the course of naive experience'' (Bernet, 1993, p. 61 - quoting Husserl). Thus, seeking objectivity in the realm of the phenomenological reifies the mind by positing the phenomenon as something independent of the experience of the phenomenon. As Merleau-Ponty poetically put it, 'reflection does not withdraw from the world towards the unity of

consciousness as the world's basis; it steps back to watch the forms of transcendence fly up like sparks from a fire; it slackens the intentional threads which attach us to the world and thus brings them to our notice' (1962, p. xiii).

Thus, the phenomenological reduction radicalizes the real significance of Churchland's eliminativism: naive assumptions gleaned from the natural attitude are not reliable. The phenomenological reduction assures that phenomenology is not simply an intellectualizing of folk psychology. Theoretically, then, the question is not about the truth value or the usefulness of any given naive assumption. Whatever those assumptions are, we must find a way to get behind them and subject them to examination on grounds that are universally accessible.

More importantly, the phenomenological reduction is a methodological tool essential to phenomenology in the Husserlian tradition. It is the only point of access to universal truth regarding phenomenology's subject matter. If we are to explain phenomenal experiences ('qualia' in the jargon of analytic philosophy) then we must give them a dignity appropriate to the subject matter. For phenomenology, which takes as its subject matter experience itself, the only useful approach is to approach experience as it is found in nature (which is always in the first person) and to explain it as well as possible. Thus, the question of whether phenomenology is possible is the question of whether the phenomenological reduction is possible. Unless we apprehend the universal through that which is given to us in immediate experience, phenomenological investigation is solipsistic.

Is a phenomenological reduction possible? Nagel suggests his speculative proposal would eventually lead to a brick wall, and no one in analytic philosophy has

taken it up. Even Merleau-Ponty recognizes that one lesson of the reduction is that a complete reduction is not possible. So, the issue is not one of complete reduction but of a complete enough reduction to say something universal about experience. I would guess that Churchland believes this it is not possible. Fortunately, I do not need to claim that she must accept the possibility of phenomenological reduction, only that she could.

She might agree that there is a realm of investigation, the phenomenological, which examines subjective experience as a natural phenomenon and is concerned with it in its universality. Phenomenological reduction does preclude the possibility of an ontological simplification, but not of an intertheoretic interconnection. That is, phenomenological investigations can not accurately be described as ‘nothing but the neural event’ (or social fact, etc.) but descriptions of the neural events which give rise to phenomenological investigation are not irrelevant. Thus, the three-tiered approach involves a version of physicalism that does not ontologically simplify.<sup>11</sup>

Phenomenological investigations might be described as ‘constituted by, but not reducible to’ neural phenomenon ‘x’. And the account of this neural phenomenon must be prepared for transcendence. They must in some way capture how the neural events relate in such a way as to transcend mere determined biological happening and reach out towards the world. There is nothing about Churchland’s approach which requires her, *a priori*, to reject the possibility of such mechanisms. And her eliminativism suggests that she ought to be open minded. After all, she too is in favor of a sort of bracketing of the

---

11. Interestingly, whereas Churchland suggests that ontological simplification is a characteristic feature of reductive explanation, Ernest Nagel is clear that reductive explanations need not imply ontological simplification. As Nagel puts it ‘a term in a reduced law may be a predicate which refers to some distinctive *attribute* or characteristic of things. [That is,] the bridge law may specify the conditions...under which the attribute occurs’ (105).

natural attitude. She ought to be sympathetic to attempts to look upon subjective experiences as they are given.

### **III. Plasticity: Application of Three-Tiered Model**

The above arguments can only demonstrate that there is no *a priori* reason that Churchland or Merleau-Ponty would reject a three-tiered approach; such an explanatory framework is in principle possible. Many (including Churchland, I would guess) would argue, *as a matter of empirical reality*, that phenomenology does not meet the constraints which emerge ultimately from neuroscience. Many (including Merleau-Ponty, I would guess) would argue, *as a matter of empirical reality*, that neuroscience has nothing interesting to contribute to phenomenology. Making the case that collaboration between Merleau-Ponty and Patricia Churchland is logically possible is not yet to make the case that such collaboration is of any practical benefit.

The following discussion of neural plasticity is intended to serve two purposes. First, to demonstrate that explanatory unity across three levels of analysis already exists. That is, the three-tiered model is consistent with existing research across phenomenology, psychology, and neuroscience in that all three levels of investigation understand the mind-brain to be plastic. Second, to speculate about the practical benefits of collaboration between neuroscience and phenomenology regarding our understanding of plasticity. That is, to show that future research on plasticity in both neuroscience and phenomenology might benefit from a three-tiered explanatory model.

#### *A. Three tiers of analysis*

While it is natural to hold that the ultimate standard for evaluating a theoretical model is whether it can lead to *future* explanations, initial evaluations of a theoretical model must be based on the model's ability to account for *existing* research. As pointed out earlier, neuroscience and phenomenology have developed thus far along independent trajectories. If it can be shown that those independent trajectories have resulted in claims that can be coherently unified, a *prima facie* case will have been made that the three-tiered model reflects the actual structure of conscious experience.

--*Plasticity: neuroscience, psychology, and phenomenology*

There is much recent work on neural plasticity, the ability of a neuron (or network of neurons) to reshape itself both physically and functionally in response to various kinds of stimulation. For the sake of brevity, I will mention only a study by Melchner et al. which provides a stunning demonstration of the plasticity of neurons through experiments on ferrets. Their procedure was to examine 'ferrets in which retinal projections are redirected neonatally to the auditory thalamus' (2000, p. 871). Within the first day after the ferret's birth, surgery was performed to remove the primary visual cortex and reroute visual stimulus to the auditory cortex. Through a variety of experimental techniques, they were able to determine that the ferret's auditory thalamus was able to support visual experience. Specifically, the auditory cortex was able to support localization of visual stimulus with an orientation map. That is, the ferrets were able to determine the relative spatial locations of various stimuli. Thus, the visual orientation map is not, as previously thought, an intrinsic feature of the primary visual cortex; even V1 is actively constructed in response to external stimulus. The neural tools of visual perception are not pre-

established in the biological tissues; the function of visual hardware emerges through interaction with visual stimulus.

Analysis of plasticity at the level of clinical neuropsychology is offered by studies of language acquisition and re-acquisition following left hemidecortication<sup>12</sup> (the surgical removal of one cortical hemisphere to address certain severe epileptic conditions). As Paul Broca famously observed, ‘we speak with the left hemisphere’. So, what happens when there is no left hemisphere? For many years, it has been known that if left hemidecortication happens at a young enough age, language outcomes are reasonably good and recent work is overturning the long-held belief that the isolated right hemisphere can not produce normal language (see Curtiss, de Bode, 2000). A recent case study presented Alex, who was able to develop effective if imperfect linguistic skills for the first time at the age of nine, less than a year of left hemidecortication freed him from seizures (see Vargha-Khadem et al., 1997). Given these and other findings regarding neural plasticity, surprising and substantial neural re-organization is now known to be possible – even into old age. In response to trauma, cortical regions can take on new functions. Again, the brain is shaped by visual and linguistic input just as it supports visual and linguistic output.

According to Merleau-Ponty, the body is always in communion with the world. In *Phenomenology of Perception*, he writes of a ‘coition, so to speak, of self and object’. In *The Visible and the Invisible*, he refers to a ‘reversibility of the flesh’ in which, for my relationship with anything in the world, there is a possible (at least in principle) reversing of subject and object - much as there is when I hold my own hand. Broadly considered,

---

12. More commonly, though less precisely, called ‘hemispherectomy’. It is, after all, not the entire hemisphere which is excised, but only the cortical hemisphere.



the insight is that we do not act *in* the world as much as we interact *with* the world. We constitute ourselves as embodied through this interaction and we form our ego in relation to that which we encounter. Thus, central to phenomenology is the recognition that the conscious mind is shaped through reciprocal interaction with the external environment. The mind is shaped by the so-called external world and, in turn, the mind creates the lived aspect of an external world. Thus, phenomenology rejects both the Lockean view of the mind as a blank slate upon which experience writes and the Cartesian view of the mind as an immutable substance. The lived world is actively constructed through a collaboration between self and world. The very content of qualitative, lived experience is plastic – shaped and molded in a reciprocal interplay with the world.

*--Three unified tiers*

Perhaps the unity of these three levels of analysis is already clear. At the cellular level, neuroscience shows that the task of specific brain structures is not purely given by genetics but instead is formed in relation to input. In response, those brain structures are responsible for forming such inputs into a meaningful phenomenal experience. At the clinical level, psychology shows that the brain is able to functionally reorganize in response to traumatic (if medically necessary) brain injury. Cellular re-organization is thus demonstrated at a higher-level of analysis; neural plasticity is shown to be a force that can be observed in the life of actual patients. At the global level, phenomenology shows that the mind interacts with the environment in construction of lived experiences. Across three levels of explanation, existing research suggests a coherent theoretical framework in which phenomenology gives insight into the lived significance of

neuroscientific findings and neuroscience gives insight into physical mechanisms by which the lived world comes to be.

To think about the relationship between intersubjectivity and neural plasticity in this way is not to ontologically simplify intersubjectivity into neural plasticity; ‘intersubjectivity’ and ‘neural plasticity’ are not two terms for one thing and there are no bridge laws available (or likely to be ever found) that can define intersubjectivity in neural terms. But the three-tiered model does not suggest that there need be. Phenomenology, like psychology and neuroscience, is an in-eliminable field of investigation for which the lived experience as such is its investigative domain. ‘Reduction’ is the right word to describe the research agenda in the sense that observations at a lower level of analysis explain why the higher level facts are what they are. Similarly for ‘eliminativism’. Thinking about plasticity on the three-tiered model does not eliminate intersubjectivity in favor of a neuroscientific account. Nevertheless, phenomenology has itself largely eliminated the subject-object distinction in its bracketing of the natural attitude. Instead of eliminating phenomenology, a three-tiered explanatory model grounds phenomenological accounts of intersubjectivity in the broader context of scientific knowledge.

It is no objection to point out that the phenomenology and neuroscience relevant to plasticity have developed independently. Far from supporting the claim that interaction between neuroscience and phenomenology is unnecessary, the existing explanatory cohesion provides the necessary points of contact for fruitful interaction.

#### *B. Utility of a co-evolution approach*

Before speculating about future research, I want to point out that the recent explosion of literature on neuroscience and phenomenology can be understood precisely in terms of the three-tiered model. Some point to neuroscientific findings as confirmation of phenomenological insights. Some pursue the neural mechanism of phenomenological observations (Gallagher 2005). Others turn to phenomenology for deeper insight into clinical observations about, for instance, neglect. Explanatory patterns, conceptual refinement, and corroborating evidence are increasingly being sought across levels of investigation. In light of so much existing research that can be understood in a three-tiered model, my brief speculations will address what are probably the two questions which Merleau-Ponty and Churchland, respectively, would likely be most skeptical about: ‘Can neuroscience have a corrective influence on phenomenology?’ and ‘Can collaboration with phenomenology payoff in terms of actual neuroscientific research results?’.

*--Neuroscience constraining phenomenology*

One reason the previously cited studies by Melchner et. al is such a wonderful example to use in explaining the existing explanatory coherence between phenomenology and neuroscience is that it calls into question many long held beliefs about the limits of neural plasticity. Nevertheless, the neuroscientific and psychological literature compelling makes the case that there are limits to neural plasticity. The case of Alex, who recovered speech at nine years old, is a striking example of the capacities of the human brain precisely because Alex is the exception. After left-hemidecortication, language function is generally damaged to some degree; many have argued that the right

hemisphere is incapable of supporting certain syntactical complexities. Even in the studies by Melchner et. al, there is no suggestion that the auditory cortex of the ferrets could process visual stimulus with the same strength or complexity as V1 could.

Too often phenomenological descriptions seem likely to be overstating the inherent capacity of individuals to constitute (or to have constituted) the lived world differently. For example, Sartre surely goes too far in saying that for human beings 'existence precedes essence'. Were the brain limitlessly plastic, Sartre would be right; if our perceptions, linguistic conventions, emotional states and so on could be limitlessly reshaped, our essence would indeed be determined by the unfolding of our existence. The fact that the brain is substantially plastic confirms that Sartre is partially right. An understanding of the cellular limitations on plasticity could correct phenomenological descriptions which represent the conscious being as limitlessly pluri-potential in its adaptive interactions with the so-called objective and external world. Similar analysis can be applied to Merleau-Ponty's previously cited claim about the 'coition of self and object'. A understanding of the extent of the limits of neural plasticity, on the cellular level, might lead phenomenologists to temper their claims about the extent to which the self is formed in relation to external world. Understanding the nature of those limits might provide phenomenologists with subtle insights into why certain aspects of lived experience are given.

*--Phenomenology practically beneficial to neuroscience*

For Merleau-Ponty (and Husserl) plasticity is manifest not on the level of individual neurons or cortical regions; it is as a whole and unified conscious being that

one interacts with the world. Thus, two closely related pieces of advice that phenomenology would give to neuroscience are particularly relevant to plasticity. First, phenomenology would advise neuroscience to reclaim the body. Merleau-Ponty writes,

‘For contemporary psychology and psychopathology the body is no longer merely an object in the world, under the purview of a separated spirit. It is on the side of the subject; it is our point of view on the world, the place where the spirit takes on a certain physical and historical situation. As Descartes once said profoundly, the soul is not merely in the body like a pilot in his ship; it is wholly intermingled with the body. The body in turn is wholly animated, and all its functions contribute to the perception of objects’ (1964b, p. 5).

The mind can not be understood if it is examined outside of its relation with the entire body. Seeing an object, for instance, can not be understood in its full phenomenal color simply by examining the wave lengths affecting rods and cones and then mapping out where these signals are sent, only to end up in detailed examinations of the brain processing these ‘signals’. The whole of the perceptual apparatus must be brought to bear. On the contrary, too often neuroscience removes the brain from the body when studying it, cutting it just below the brain stem (both literally and figuratively). Though there has been extensive examination of the workings of the peripheral nervous system, even these studies persist in the separation of the central and peripheral nervous systems. And the peripheral nervous system is still understood as a passive-receptive system which simply takes in certain ‘stimulus’ from the environment and sends signals to the brain.

Neuroscience should focus more on the nature of peripheral inputs, so as to better understand how they inform cortical processing. After all, a neuroscientific correlate for the embodiment of the subject must provide an understanding of the peripheral nervous system's impact on higher order processing such as reflection, memory, language etc.

Research into language reacquisition after left hemidecortication might focus more on the construction of symbolic meaning through interaction with the world than on the presence or absence of certain complex syntactical expressions. There might be better understanding of the relationship between linguistic functioning and perceptual motor skills. Maybe it will turn out that the abrupt right visual field cut and right hemi-paresis creates perceptual motor challenges that can inhibit linguistic recovery. If so, there would certainly be clinical applications regarding appropriate physical, occupation, and speech therapy regimens following left hemidecortication that vary depending on factors such as precise surgical technique, brain imaging results and so on. On a cellular level, perhaps it will turn out that the peripheral nervous system is itself dynamic and interactive. Perhaps it will turn out that brain function and organization can not be explained without comprehension of the structural coherence of cortical functioning with the peripheral nervous system. Perhaps there will even be a recognition that the neural processes constituting conscious experience necessarily include an active contribution from peripheral nerve cells.

Of course, investigation of the peripheral nervous system and its relationship with the central nervous system can only get us so far in providing a neural basis for the embodiment of consciousness. After all, for Merleau-Ponty, consciousness is not abstract embodiment in a particular biological organism. For Merleau-Ponty, embodiment places

the subject within the world, in a particular social, historical, cultural and spatial context. Thus, the second piece of phenomenological advice for neuroscience: locate the mind in its social and cultural environment.

Neuroscience must move not only from the brain to the central nervous system, but from the central nervous system out into the world. We are always already in a social situation and nothing we do amounts to an escape from it. We have no access to things-in-themselves, independent of our perceptions of them. Nevertheless, starting from the lived-world does not condemn us to solipsism. We are able to experience a shared world on the basis of intersubjectivity. Thus, essential to the mind is its relationship to the other. Therefore, it is not enough for neuroscience to stand in intertheoretic relationship with psychology. Sociology, anthropology and history must be integrated with psychology, and neuroscience must reflect this integration. Only on this basis do we see the real extent of the mistake of cutting the brain at the bottom of the brain stem and building a theory of the mind-brain on study of what is left.

It is not difficult to see how such expansion of orientation might shape research into neural plasticity on both the psychological and neuroscientific level. On the psychological level, it would perhaps emerge that language re-acquisition after left hemidecortication would be enhanced were patients to collaborate in speech therapy. Perhaps it will turn out that language can re-emerge more completely were post-operative speech therapy to focus more on the social and cultural touchstones of the patient. More generally, post-operative recovery would benefit from more robust attempts to empathetically understand the lived experience of left hemidecortication and thus connect the brain damaged patient more coherently with the patient's social and cultural milieu.

#### **IV. Conclusion**

Any theory of the brain which takes phenomenology seriously must see the brain as a dynamic, changing, interactive, malleable and engaged organ which encounters, shapes and ‘communes’ with its environment. Not a passive receiver of external contact, but an organ actively engaged in the construction of perception. Not a static given which merely shapes its perceptions on the basis of its nature, but a changeable and changing entity which constructs itself in making various sensations into a perceived world. In general, we need a brain which does not merely act on the world, but interacts with it. Of course, there are hard questions that I have hardly touched on. Can neuroscience recognize the existence of the Ego, of intentionality, of reflection? Would there need to be a neurobiology of phenomenological reduction? What would such a thing possibly look like? Given that both Churchland and Merleau-Ponty are highly critical of *existing* psychology, would psychology be able to bear the weight applied on it from both above and below?

Though this defense of the three-tiered model is limited in scope, I hope I have at least made the case that analytic physicalism and European phenomenology can fruitfully interact. Perhaps someday the currently shrinking divide between analytic and continental philosophies of mind will vanish entirely and a single coherent conversation (including philosophers as well as psychologists and neuroscientists) about subjectivity,



intentionality, and perception can develop. Then we might have an explanation of consciousness.<sup>13</sup>

---

<sup>13</sup> Many thanks to two anonymous *Journal of Consciousness Studies* referees for insightful critiques of an earlier draft. Thanks also to my hemidecorticated son Ciaran for inspiring my interest in phenomenal consciousness.

## Bibliography

- Alter, Torin. "A Limited Defense of the Knowledge Argument," *Philosophical Studies* 90 (1998), 35-56.
- Bickle, John and Ellis, Ralph. "Phenomenology and Cortical Microstimulation." In *Phenomenology and Philosophy of Mind*, edited by David Woodruff, 140-166. Oxford: Clarendon Press, 2005.
- Bigelow, John and Pargetter, Robert. "Acquaintance with Qualia." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 179-196. Cambridge: MIT Press, 2004. Originally published in *Theoria* 61 (1990): 129-147.
- Block, Ned and Stalnaker, Robert. "Conceptual Analysis, Dualism, and the Explanatory Gap." *The Philosophical Review* 108 (1999): 1-46.
- Block, "Mental Paint." In *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, edited by Martin Hahn and Bjorn Ramberg. Cambridge: MIT Press, 2003.
- Boyle, Noel. "Neurobiology and Phenomenology: Towards a Three-Tiered Intertheoretic Model of Explanation." *Journal of Consciousness Studies* 15, no. 3 (2008): 34-58.
- Brunet, Rudolf et al., *An Introduction to Husserlian Phenomenology*. Evanston, IL: Northwestern University Press, 1993.
- Carr, David. "Phenomenology and Fiction in Dennett," *International Journal of Philosophical Studies*, 6 (1998): 331-44.
- Chalmers, David. "Facing Up to the Problem of Consciousness," *Journal of Consciousness Studies*, 2 (1995): 200-19.
- Chalmers, David. *The Conscious Mind*. Oxford University Press, 1996.
- Chalmers, David. "Phenomenal Concepts and the Knowledge Argument." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 269-298. Cambridge: MIT Press, 2004.
- Chalmers, David. "Epistemic Two-Dimensional Semantics." *Philosophical Studies* 118 (2004): 153-226.
- Churchland, Patricia. *Neurophilosophy*. Cambridge: MIT Press, 1986.

- Churchland, Patricia. "Do We Propose to Eliminate Consciousness?" In *The Churchlands and Their Critics*, edited by Robert McCauley, 297-300. Cambridge: Blackwell Publishers, 1996.
- Churchland, Patricia. "Can neurobiology teach us anything about consciousness?" In *The Nature of Consciousness*, edited by Ned Block et al., 127-141. Cambridge: MIT Press, 1997. Originally published in *Proceedings and Addresses of the APA* 67 (1995): 23-40.
- Churchland, Paul. "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* 82 (1981): 67-90.
- Churchland, Paul. "Reduction, Qualia, and the Direct Introspection of Brain States." *Journal of Philosophy* 94 (1985): 8-28.
- Churchland, Paul. "Knowing Qualia: A Reply to Jackson." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 163-178. Cambridge: MIT Press, 2004. Originally published in Paul Churchland. *A Neurocomputational Perspective*, 67-76. Cambridge: MIT Press, 1989.
- Churchland, Paul and Churchland, Patricia. "Intertheoretic reduction: A neuroscientist's field guide." *Seminars in the Neurosciences* 4 (1990): 249-256.
- Conee, Earl. "Phenomenal Knowledge." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 197-216. Cambridge: MIT Press, 2004. Originally published in *Australasian Journal of Philosophy* 72 (1994): 136-150.
- Crick, Francis and Koch, Christopher. "Towards a Neurobiological Theory of Consciousness." In *The Nature of Consciousness*, edited by Ned Block et al., 277-292. Cambridge: MIT Press, 1997. Originally published in *Seminars in the Neurosciences* 2 (1990): 263-275.
- Curtis, Susan, de Bode, Susan. "How normal is grammatical development in the right hemisphere following hemispherectomy?" *Brain and Language* 86 (2003): 193-206.
- Decety, J. et. al. "Mapping Motor Representation with Positron Emission Tomography." *Nature* 371 (2002): 600-602.
- Dennett, Daniel. *Consciousness Explained*. Boston: Little, Brown Publishers, 1991.
- Dennett, Daniel. "Epiphenomenal Qualia?" In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 59-68. Cambridge, MIT

- Press, 2004. Originally published in Daniel Dennett, *Consciousness Explained*, Boston: Little, Brown Publishers, 1991.
- Gallagher, Shawn. *How the Body Shapes the Mind*. Oxford: Clarendon Press, 2005.
- Graham, Todd. "Conceptualizing the (Dis)unity of Science." *Philosophy of Science* 71 (2004): 133-55.
- Graham, George and Horgan, Terence. "Mary, Mary, Quite Contrary." *Philosophical Studies* 99 (2000): 59-87.
- Hardcastle, Valerie. "The Why of Consciousness: A Non-Issue for Materialists." *Journal of Consciousness Studies* 1 (1996): 7-13.
- Hardcastle, Valerie. "Reduction, Explanatory Extension, and the Mind/ Brain Sciences." *Philosophy of Science* 59 (1992): 408-428.
- Hellie, Benj. "Inexpressible Truths and the Allure of the Knowledge Argument." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 333-364. Cambridge: MIT Press, 2004.
- Hempel, Carl. *Aspects of Scientific Explanation*. New York: The Free Press, 1965.
- Hempel, "Two Basic Types of Explanation." In *Philosophy of Science*, edited by Martin Curd and J.A. Cover, 685-694. London: W.W. Norton & Company, 1998. Originally published as "Explanation in Science and History." In *Frontiers of Science and Philosophy*, edited by R.G. Colodny, 9-19, 32. London and Pittsburgh: Allen and Unwin and University of Pittsburgh Press, 1962.
- Horgan, Terence. "Jackson on Physical Information and Qualia." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 301-308. Cambridge: MIT Press, 2004. Originally published in *The Philosophical Quarterly* 135 (1984): 147-152.
- Husserl, Edmund. *Cartesian Meditations*, translated by Dorion Cairns. The Netherlands: Dordrecht, 1995.
- Jackson, Frank. "Epiphenomenal Qualia." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 39-50. Cambridge: MIT Press, 2004. Originally published in *Philosophical Quarterly* 32 (1982): 127-136.
- Jackson, Frank. "What Mary Didn't Know." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 51-56. Cambridge: MIT

Press, 2004. Originally published in *The Journal of Philosophy* 83 (1986): 291-295

Jackson, Frank. "Armchair Metaphysics." In *Philosophy in Mind*, edited by John O'Leary Hawthorne and Michaelis Michael, 23-42. Dordrecht: Kluwer, 1994.

Jackson, Frank. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford University Press, 1998.

Jackson, Frank. "Mind and Illusion." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 421-442. Cambridge: MIT Press, 2004.

Jackson, Frank. "A Priori Physicalism." In *Contemporary Debates in Philosophy of Mind*, ed. Jonathan Cohen and Brian McLaughlin, 185-199. Basil Blackwell, 2007. (When I read the article it was still forthcoming. Frank Jackson attached it to an e-mail in which he kindly answered a question about his current position.)

Jackson, Frank and Chalmers, David. "Conceptual Analysis and Reductive Explanation." *The Philosophical Review* 110, no. 3 (2001): 315-360.

Jones, Todd. "Reductionism and the Unification Theory of Explanation." *Philosophy of Science* 62 (1995): 21-30.

Kelly, Sean. *The Relevance of Phenomenology to the Philosophy of Language and Mind*. New York: Garland Publishing, 2001.

Kitcher, Philip. "1953 and all That: A Tale of Two Sciences." *The Philosophical Review* 93, no. 3 (1984): 335-373.

Kitcher, Philip. "Explanatory Unification and Causal Structure." In Philip Kitcher and Wesley Salmon, editors, *Scientific Explanation: Minnesota Studies in the Philosophy of Science*, 410-506. University Minnesota Press, 1989.

Kolb, Bryan and Whishaw, Ian. *Fundamentals of Human Neuropsychology*, fifth edition. New York: Worth Publishers, 2003.

Kripke, Saul. *Naming and Necessity*. Cambridge: Harvard University Press, 1972.

Levine, "Materialism and Qualia." *Philosophical Quarterly* 64 (1983): 354-361.

Levine, "Materialism and Qualia." *Pacific Philosophical Quarterly* 64 (1983): 354-361.

David Lewis. "What Experience Teaches." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 77-104. Cambridge, MIT

- Press, 2004. Originally published in *Proceedings of the Russellian Society* 13 (1998): 29-57.
- Brian Loar. "Phenomenal States (Revised Version)." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 219-240. Originally published in Ned Block, Owen Flanagan, and Guven Guzeldere, editors, *The Nature of Consciousness*, 597-616. Cambridge: MIT Press, 2004.
- McCauley, Robert. "Explanatory Pluralism and the Co-Evolution of Theories in Science." In *The Churchlands and Their Critics*, edited by Robert McCauley, 1-16. Cambridge: Blackwell Publishers, 1997.
- McGinn, Colin. *The Mysterious Flame*. New York: Basic Books, 1999.
- Melchner et al. "Visual behavior mediated by retinal projections directed to the auditory pathway." *Nature* 404 (2000), 871-875.
- Merleau-Ponty, Maurice. *The Phenomenology of Perception*, translated by Colin Smith. London: Routledge, 1962.
- Merleau-Ponty, Maurice. *The Visible and the Invisible*, translated by Alphonso Lingis. Evanston, IL: Northwestern University Press, 1968.
- Merleau-Ponty, Maurice. "Phenomenology and the Sciences of Man." In *The Primacy of Perception*, translated by John Wild, 43-95. Evanston IL: Northwestern University Press, 1964.
- Merleau-Ponty, Maurice. "An Unpublished Text by Maurice Merleau-Ponty: A Prospectus of His Work", translated by Arleen Dallery. In *The Primacy of Perception* 3-11. Evanston, IL: Northwestern University Press, 1964.
- Merleau-Ponty, Maurice. "Sociology and the Philosopher," translated by Richard McCleary. In *Signs*, 98-113. Evanston, IL: Northwestern University Press 1964.
- Musacchio, Jose. "Dissolving the Explanatory Gap: Neurobiological Differences Between Phenomenal and Propositional Knowledge." *Brain and Mind* 3 (2002): 331-365.
- Nagel, Ernst. "Issues in Reduction." In *Philosophy of Science*, edited by Martin Curd and J.A. Cover, 905-921. London: W.W. Norton & Company, 1998. Originally published in Ernst Nagel, *Teleology Revisited*, 95-13. New York: Columbia University Press, 1974.
- Nagel, Thomas. "What is it like to be a Bat?" *Philosophical Review* 4 (1976): 435-450.

- Lawrence Nemirow. "Review of *Mortal Questions* by Thomas Nagel." *Philosophical Review* 89 (1980): 473-477.
- Overgaard, Morten. "On the Naturalizing of Phenomenology." *Phenomenology and the Cognitive Sciences* 3 (2004): 365-79.
- Thomas Polger. *Natural Minds*. Cambridge: MIT Press, 2004.
- Karl Popper. "Science: Conjectures and Refutations." In *Philosophy of Science*, edited by Martin Curd and J.A. Cover, 3-10. London: W.W. Norton & Company, 1998. Originally published in Karl Popper, *Conjectures and Refutations*. London: Routledge and Kegan Paul, 1963.
- Rowling, J.K. *Harry Potter and the Deathly Hallows*. New York: Arthur A. Levine Books, an imprint of Scholastic Press, 2007.
- Rumelhart, D., McClelland, J.. "On Learning the Past Tense of English Verbs." In *Parallel Distributed Processing: explorations in the microstructure of cognition*, vol. 2, chapter 18. Cambridge, MA: MIT Press, 1986.
- Sacks, Oliver. *An Anthropologist on Mars*. New York: Vintage Books, 1995.
- Salmon, Wesley. "Scientific Explanation: Causation and Explanation." In *Causality and Explanation*, 68-78. New York: Oxford University Press, 1998. Originally published in *Critica* 22 (1990): 3-21.
- Salmon, Wesley. "The Importance of Scientific Understanding." In Wesley Salmon, *Causality and Explanation*, 79-92. New York: Oxford University Press, 1998.
- Salmon, Wesley. "A New Look at Causality." In Wesley Salmon, *Causality and Explanation*, 13-24. New York: Oxford University Press, 1998.
- Salmon, Wesley. "Dreams of a Famous Physicist." In Wesley Salmon, *Causality and Explanation*, 385-404. New York: Oxford University Press, 1998.
- Salmon, Wesley. "A Third Dogma of Empiricism." In Wesley Salmon, *Causality and Explanation*, 95-107. New York: Oxford University Press, 1998. Originally published in *Basic Problems in Methodology and Linguistics*, edited by R. Butts and J. Hintikka, 149-166. Dordrecht: D. Reidel, 1977.
- Salmon, Wesley. "An 'At-At' Theory of Causal Influence." In Wesley Salmon, *Causality and Explanation*, 193-199. New York: Oxford University Press, 1998. Originally published in *Philosophy of Science* 44 (1977): 215-224.

- Stawarska, Beata. "Merleau-Ponty in Dialogue with the Cognitive Sciences in Light of Recent Imitation Research." *Philosophy Today* 47 (2003): 88-99.
- Steel, Dan. "Can a reductionist be a pluralist?" *Biology and Philosophy* 19 (2004): 55-73.
- Stoljar, Daniel. "Two Conceptions of the Physical." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 309-332. Cambridge, MIT Press, 2004. Excerpted from "Two Conceptions of the Physical," *Philosophy and Phenomenological Research*, (2001): 253-270.
- Tye, Michael. "Knowing What it is Like." In *There's Something About Mary*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar, 143-160. Cambridge, MIT Press, 2004. Originally published in Michael Tye, *Consciousness, Color, and Content*, 3-20. Cambridge: MIT Press, 2000.
- Weinberg, Steven. *Dreams of a Final Theory*. New York: Vintage Books, 1992.
- Vargha-Khadem et al. "Onset of speech after left hemispherectomy in a nine-year-old boy." *Brain* 120 (1997): 159-182.
- Wilder, Kathleen. *Bodily Nature of Consciousness: Sartre and Contemporary Philosophy of Mind*. Ithaca: Cornell University Press, 1997.



MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03062 4856