Fall 2010

A defense of the knowledge argument

John Martin DePoe
University of Iowa

A DEFENSE OF THE KNOWLEDGE ARGUMENT

by

John Martin DePoe

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy degree
in Philosophy
in the Graduate College of The University of Iowa

December 2010

Thesis Supervisor: Professor Richard A. Fumerton

ABSTRACT

Defenders of the Knowledge Argument contend that physicalism is false because knowing all the physical truths is not sufficient to know all the truths about the world. In particular, proponents of the Knowledge Argument claim that physicalism is false because the truths about the character of conscious experience are not knowable from the complete set of physical truths. This dissertation is a defense of the Knowledge Argument. Chapter one characterizes what physicalism is and provides support for the claim that if knowing all the physical truths is not sufficient to know all the truths about the world, then physicalism is false. In chapter two, I defend the claim that knowing all the physical truths is not sufficient for knowing all the truths about the world. In addition to mounting a *prima facie* case for the knowledge intuition, I present and defend an epistemology grounded in direct acquaintance to provide a more substantive argument to accept it.

Chapters three through five address the physicalist objections to the Knowledge Argument. The first set of objections advocates that knowing all the physical truths is, in fact, sufficient for knowing all the truths about the world. The next set of objections admits that there is some sense in which knowing all the physical truths is not sufficient for knowing all the truths about the world. However, these objections maintain that the kind of knowledge that is absent from the complete set of physical truths is know-how or knowledge by acquaintance, and not factual or propositional knowledge. The final set of

objections maintain that the kind of propositional knowledge that is left out of the complete set of physical truths is compatible with physicalism. My response to these objections is part of advancing my *prima facie* case for the Knowledge Argument.

The final chapter addresses a structural question that pertains to the Knowledge Argument. Some philosophers have maintained that the structure of the Knowledge Argument invites a kind of self-refutation of any systematic account of reality. The concern is that the Knowledge Argument proves too much, and that the dualist who uses the argument to refute physicalism risks the argument defeating his own position. I will argue that the Knowledge Argument does not refute dualism.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

A DEFENSE OF THE KNOWLEDGE ARGUMENT

by

John Martin DePoe

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy degree
in Philosophy
in the Graduate College of The University of Iowa

December 2010

Thesis Supervisor: Professor Richard A. Fumerton

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph. D. thesis of

John Martin DePoe

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Philosophy at the December 2010 graduation.

Thesis Committee: _____
Richard A. Fumerton, Thesis Supervisor

_____
David Cunning

_____
Evan Fales

_____
Ali Hasan

_____
Frederick Skiff

To
Jeannie, my beloved wife,
and
Mary, my favorite scientist,
both of whom have suffered much to see this project through.

ACKNOWLEDGEMENTS

Space will not permit a thorough recognition of everyone who has helped me in some way or another in the process of preparing and finishing this project. However, I would like to acknowledge a few individuals who have helped to see me through this project. First, I would like to thank my wife, Jeannie, who has put up with this year-long project. Without her full support, encouragement, friendship, and love, I may not have finished this work so quickly and without losing my sanity.

I would be remiss if I did not thank Richard Fumerton for his guidance that saw me through this dissertation. It was through taking Richard's course on the Knowledge Argument that I took interest in this topic and began to conceive of the basic structure of thought that ultimately became realized in the pages that follow. Richard always seemed to make time for me and my half-baked ideas, even when he was obviously busy with his own work and responsibilities.

Without encouragement I would not have pursued higher education. So, I'd like to thank my parents, Scot Miller, Dan Stiver, Timothy McGrew, John Dilworth, and the University of Iowa's philosophy faculty for encouraging me to reach this goal. Finally, I thank the graduate college for awarding me the Seashore-Ballard Fellowship, which made it possible for me to finish this dissertation within a year as well as the philosophy department for supporting me throughout the course of this degree.

TABLE OF CONTENTS

CHAPTER 1: THE DISTINCTION BETWEEN PHYSICALISM AND DUALISM

Consciousness is considered by many to be at the heart of many genuinely hard problems in philosophy.[1] Consciousness, which is probably best understood as the subjective character of experience or what it's like to have mental experience, is one of the primary obstructions to giving a thoroughgoing physicalist account of human nature. Jaegwon Kim blames consciousness for the seemingly intractable problem of mental causation and the inability to provide a complete functional reduction of the mind.[2] Many philosophers aren't quite sure what to do with consciousness. The existence of consciousness seems undeniable unless one is making a joke,[3] but recognizing consciousness can generate seemingly insuperable problems. Other philosophers have suggested that consciousness is fundamentally the sort of phenomena that must remain a mystery and cannot be explained.[4] Most of these problems and reactions stem from the apparent incompatibility of consciousness and physicalism. Rather than lament this observation, this dissertation embraces the incompatibility of consciousness and physicalism. Indeed, one goal of this project is to show that consciousness cannot be squared with physicalism, and that the lesson to take from this is that physicalism is false.

---

[1] See, for example, Nagel (1974), Chalmers (1995a), and Kim (2005).

[2] Kim (2005).

[3] Although some philosophers seem to have made this claim with a straight face, such as Rorty (1965), Rorty (1970), Stich (1983), and Churchland (1995).

[4] Such as Nagel (1986) and McGinn (1999).

The central argument that I will press to justify the conclusion that physicalism is false is commonly called "The Knowledge Argument," which contends that our knowledge of certain conscious experiences are incompatible with physicalism. In order to make good on this argument, I will proceed to defend the argument in the following progressive steps. In the first chapter, I will provide an account of the central concepts that are employed in this debate, such as physicalism and dualism. The second chapter will lay the epistemological groundwork for the Knowledge Argument by showing that direct acquaintance secures our knowledge of conscious experience. In essence, the first two chapters present a positive case for the Knowledge Argument. The negative case for the Knowledge Argument is taken up in chapters three through five, which embark on showing that various physicalist responses to the Knowledge Argument cannot adequately show the compatibility of conscious experience and physicalism. The final chapter responds to the objection that states that the Knowledge Argument can be applied to dualism, thereby revealing that it proves too much.

There is a lot of ground to cover in order to argue for my controversial thesis. This chapter is the first step on this long journey. In this chapter I will give an account of physicalism and dualism. I will be arguing that the most plausible form of physicalism is committed to a reductive theory—one that if it were true, knowing the fundamental physical substances and properties would be sufficient to know all the properties and substances in the world. It will be

evident in later stages of the argument that this characterization of physicalism is crucial for justifying a key premise in the Knowledge Argument. Presently, I shall turn to this important part of my argument and show that physicalism is best understood as a reductive and *a priori* thesis.

1.1 Physicalism

Physicalism is a metaphysical thesis about the ontology of world, which in its most trivial form states that everything that exists is physical. There are, however, many different ways to cash out the notion that everything that exists is physical. Clear examples of physicalist approaches to the mind-body problem include logical behaviorism,[5] brain state identity theories,[6] and eliminative materialism.[7] All of these approaches share a feature that is essential to any account of physicalism, which is the affirmation that everything is physical in nature. By implication, this means that nothing exists that cannot be accounted for in terms of the physical. They also agree that there are no truths about the world that would be left out of a complete physical description of the world. So, physicalism is a putative account of reality where everything that exists is physical or the consequence of the relations and properties of physical things.

So, what does it mean for something to be physical? Providing a meaningful characterization of the physical is a tricky task because of Hempel's

---

[5] Such as Ryle (1949).

[6] Such as Place (1956) and Smart (1959).

[7] Such as Rorty (1965), Rorty (1970), Stich (1983), and Churchland (1995).

dilemma.[8] The dilemma states that, on the one hand, if the physical is defined in terms of what contemporary physics picks out, then physicalism is bound to be false (since contemporary physics is incomplete and probably contradictory since it affirms both quantum theory and string theory). On the other hand, if the physical is defined in terms of what an idealized, completed physics picks out, then physicalism appears to be a vague, unsubstantial thesis. These appear to be the only two options for defining the physical, and neither seems very promising. In an effort to give physicalism the most charitable interpretation, I think it is best to define the physical in terms of those things that would be directly described by an idealized, completed physics.[9] To stave off charges of being too vague, however, I think it is safe to say that we have some idea of the properties that we expect to be included and excluded from an idealized, completed physics. For example, we expect mass and charge to be included in a completed physics and irreducible qualitative mental states and Cartesian souls to be excluded. For the issues that this dissertation is engaging (e.g., the mind-body problem), we have a general idea of the sorts of properties that we expect a completed, idealized neuroscience to ascribe to the brain and what sorts of

[8] Hempel (1970). See Crane and Mellor (1990) for an argument that uses Hempel's dilemma to argue that physicalism cannot avoid the problem of triviality.

[9] I understand physics as a discipline that is defined by its methods. Such methods would include publicly observable empirical tests (among other things) and would *not* include the philosopher's methods of introspection, armchair concept analysis, and pure thought experiments.

properties that a completed, idealized neuroscience will not include among the most fundamental properties of the brain.[10]

Although there are numerous ways to be a physicalist, currently most of the paradigm physicalisms are generally considered to be untenable. For example, logical behaviorism is widely regarded to be indefensible because the translations that were supposed to hold between behavior states and mental states are generally acknowledged not to hold.[11] Brain state identity theories, which identify mental states with brain states, suffer from a number of well-known problems, perhaps most notably the problem of multiple realization.[12] Stated crudely, the problem is that brain states cannot be identical with mental states because the same mental states can be realized by other kinds of physical states. Since both humans and octopi realize pain states without being in the same brain states, for example, it follows that pain cannot be identical with a particular type of brain state. Likewise, eliminative materialism seems untenable in its most extreme form where it is taken to eliminate all the features of folk psychology (such as all beliefs, desires, intentions, and pains).[13] In its extreme

---

[10] This is similar to the approach taken in Maddell (1988), p. 5: "there is a notion of the physical which seems reasonably clear: what is physical is that which the physical sciences recognize to be such, and that in turn suggests a view of the universe as consisting of assemblies of elementary particles, a view which the great majority of those who call themselves materialists operate with."

[11] See, for example, Putnam (1975a).

[12] See, for example, Putnam (1967) and Fodor (1974). For an overview see Bickle (2008), §1.2.

[13] It is interesting to note that some notable eliminativists have not urged for a total elimination of folk psychology, but rather have urged for the more modest claim that we strive to

form, eliminative materialism is widely regarded to be unsustainable because it is self-defeating or manifestly false.[14]

Of course, there are other accounts of physicalism that are more plausible than logical behaviorism, the identity theory, and eliminative materialism. The task I will undertake in the following sections is to assess what is the most plausible version of physicalism. I will eventually settle on a particular brand of functionalism as the most plausible version of physicalism. Nonetheless it is instructive to survey other accounts of physicalism to consider whether they are likely to be true and whether such accounts should even count as being a kind of physicalism at all.

In order to assess these disparate accounts of physicalism, I am proposing a set of criteria for judging which account of physicalism is most promising as a plausible account of that position. In addition to the usual philosophical criteria (such as logical consistency, etc.) the criteria I will apply to the different conceptions of physicalism include:

(i) positive ontological adequacy—the alleged physicalist account includes in its physicalist ontology all paradigm physical substances, properties, and relations;

(ii) negative ontological adequacy—the alleged physicalist account excludes in its ontology all paradigm non-physical substances, properties, and relations;

---

eliminate as much of folk psychology as we can, and we'll try to reduce whatever cannot be eliminated. See Churchland (1988), pp. 48-49.

[14] For these criticisms see, for example, Baker (1987), Boghossian (1990), Boghossian (1991), Reppert (1992), Searle (1992), pp. 46-49, 58-64, and Menuge (2004).

(iii) explanatory adequacy—the alleged physicalist account provides an explanation for everything that exists in terms of the fundamental physical ontology.

The first criterion is important since a putative physicalist theory that excludes paradigm physical substances, properties, and relations would be deficient. For example, if an account of physicalism satisfies the other two criteria but also denies that the property of mass or extension is physical, or that fails to include the existence of rocks among the physical objects of the world, then it would fail to satisfy the basic requirements of a physicalist worldview. Furthermore, any account that fails to satisfy criterion (i) would be evidently false, and thereby doubly unacceptable.

The second criterion is needed to judge whether a putatively physicalist ontology is defined too broadly so that it is not compatible with the existence of Cartesian souls, irreducible qualitative mental states, and traditional theism, for example.[15] Any alleged account of physicalism that is consistent with the robust existence of these sorts of entities fails to be a genuine account of physicalism on this standard.

The last criterion is important for showing that the fundamental ontology posited in a proposed account of physicalism adequately explains (or provides the proper grounds for giving an explanation of) everything that exists.[16] In

[15] For a similar list of entities that should be ruled out by physicalism, see Haugeland (1984), pp. 6-7, Cooney (2000), pp. 3-4, and Melnyk (2003), pp. 10-11.

[16] The basic idea of providing an explanation is to be understood as using the fundamental kinds of the physical ontology (e.g., fundamental things, properties, relations, laws,

particular, I have in mind the underlying explanation for the causal and dependence relations that hold between fundamental microphysical entities and macro-objects in a physical world. Even if an account of physicalism satisfied the first two criteria, it is possible that the account would not offer any explanation as to how the fundamental physical ontology accounts for the existence of the macro-ontology. With these criteria in mind, I will survey a number of different attempts to characterize physical.

1.1.1 Early Modern Materialism

Perhaps the place to begin one's analysis of physicalism is with the concept of materialism that is often associated with certain early modern philosophers, such as Thomas Hobbes and Paul-Henri Thiry Baron d'Holbach, and discussed by René Descartes, John Locke, George Berkeley, and many others.[17] Looking back at the general approach of the early modern era, one way to analyze how they defined materialism is by finding some feature or property of a thing that would identify it as either a material or immaterial substance. For example, Descartes famously claimed that having extension is the mark of being

---

etc.) to describe the world. Such explanations may be causal, but any kind of explanation that is available through the fundamental kinds of the given ontology are acceptable. It is important not to confuse *explicability* with *determinability*. If current accounts of quantum theory prove to be correct, then some fundamental physical entities can turn out to be explained by fundamental physical posits (perhaps by statistical laws) without being determinable. For an account of explanation that I am inclined to accept, see Swinburne (2001), pp. 74-119, and (2004), chs. 2-4.

[17] See Hobbes (1660), Holbach (1770), Descartes (1641), Locke (1689), and Berkeley (1710) and (1713).

a material thing, and that having thought is the mark of being a mental thing.[18] Since minds are not extended, reasoned Descartes, they are not material substances. Berkeley characterized matter as "an inert, senseless substance, in which extension, figure, and motion do actually subsist."[19] Perhaps building on the Cartesian concept of matter, Berkeley adds to Descartes's criterion of extension that material substances are incompatible with having sensible ideas and that matter is essentially inert or static. Broadly speaking, the underlying strategy for the early moderns, then, is to define material and immaterial things by picking out properties of those things that are believed to be either essential to material substances (e.g., extended, inert, senseless) or incompatible with material substances (e.g., sensible, intrinsically potent, possessing thought).

While much of the spirit of early modern materialism remains true to contemporary physicalism, the early modern approach to characterizing materialism is beset with a number of problems worth noting. First, it is difficult to present a list of essential properties that characterize material substances that satisfies criteria (i) and (ii) without begging the question against the physicalist or the non-physicalist. For example, to claim that matter is essentially incapable of thought or sensation would beg the question against materialists, like Hobbes or Holbach, who contended that the human mind is a material substance. Even if a non-question-begging list of material and immaterial properties could be

---

[18] See especially Descartes (1641), meditation 6.

[19] Berkeley (1710), section 9.

given, there is a second problem, which is that one must also show that only material substances can have material properties and only immaterial substances can have immaterial properties. Even if having thoughts or sensations is taken to be an immaterial property, a materialist could maintain that some material substances can have the property of having sensations, for example. So, those who take a similar approach to these early modern philosophers will need to provide an additional argument that shows how possessing different kinds of properties leads to the conclusion that there are different kinds of property bearers.

A further complication for early modern materialism is found in the way contemporary science has uncovered various properties of matter that are at odds with early modern presumptions. For example, Berkeley explicitly takes matter to be intrinsically inert, although current science describes the default state of the fundamental particles of matter to be in motion.[20] A more controversial example could consider whether quantum particles and superstrings have the early modern's property of being extended. If not, then it seems the property of extension, which was the most popular attribute of matter for the early moderns, is not going to be sufficient to characterize material substances.

[20] Indeed, Newton's physics, which served as a basis for many early modern materialists, maintained that the natural state of matter is rest.

A final concern that I will raise is another version of the problematic inference noted above which is that only material substances can have material properties (and likewise for immaterial substances and properties). Above, I have highlighted some examples where material substances might take on immaterial properties. Additionally, early modern materialism does not rule out the possibility that immaterial substances could have material properties like extension. For example, some dualists have held that immaterial substances have the property of being extended.[21] Once again, this highlights that one must be careful when demarcating the material (or physical) from the immaterial (or non-physical) by using properties. In sum, the limited merits and ultimate insufficiency of the early modern account of physicalism as materialism are probably best understood as a consequence of that era's underlying views about the nature of science, as well as some unsubstantiated assumptions about the nature of properties and substances.

1.1.2 Mechanistic Cause Criterion

A more recent attempt to define physicalism does so by specifying the causal powers of physical things. Specifically, William Hasker defines the physical in terms of what can play a role in physical causation, where physical causation is defined as mechanistic causation and mechanistic causation is

[21] Many dualists have held that the mental is spatially located, which implies that is extended. See Zimmerman (2006), pp. 115-116.

defined as non-teleological proximate causation.[22] One positive aspect of this account of physicalism is that it gives a definition of physicalism that is not wedded to a specific theory of physics, and thereby it is supposed to avoid Hempel's dilemma. Hasker illustrates his account of physicalism by using a thermostat. Thermostats appear to have teleological causes and explanations for their effects (for example, they appear to turn off and on the furnace for the purpose of maintaining a specific temperature). But all of the immediate causes in a thermostat's various functions are non-teleological (for example, the thermostat turns the furnace switch on because a strip of metal became bent in a way that closed an electrical circuit due to the colder ambient air). In this way, then, thermostats are physical because all of their proximate causes are non-teleological.

On this account of physicalism, human beings are at the bottom level constituted by proximate causal processes that are non-teleological. Nonetheless, humans can exhibit teleological actions like a thermostat. All of the fundamental proximate causes are non-teleological, but the whole system can mimic teleological action. In other words, the fundamental physical causes that take place in the brain occur mechanistically, but the whole system would appear to operate towards goals, ends, and purposes.

I reject Hasker's definition of physicalism because it isn't clear that his account can satisfy criteria (i), (ii), or (iii). With regard to the first criterion, there

[22] Hasker (1999), pp. 62-64.

is the concern about physical "danglers," which would include physical entities that are caused by mechanistic physical processes, but which are causally impotent themselves. For example, the appendix is clearly the product of physical causes, but the appendix is causally powerless.[23] Since the appendix is not a proximate non-teleological cause, it would fail to be a physical thing on this view of physicalism. In short, for something to count as physical on Hasker's account (since to be physical is to be a non-teleological cause), it must be a cause, but this is too restrictive since in principle there is no reason to think that everything that is physical must be a cause.

This account fails the second criterion because it isn't clear that all non-physical things will operate according to non-teleological causation. For example, it seems entirely plausible that irreducible qualitative mental states or Cartesian souls could exist in such a way that they operate according to non-teleological proximate causes. Moreover, on Hasker's account of physicalism there is no explanation of the ontological structure and dependence relations that hold between macro-objects and microphysics. For these reasons, I must judge this account of physicalism to be inadequate for my purposes.

---

[23] One may quibble that the appendix is causally efficacious, such as when we observe the appendix, then, it is causally responsible (in no small way!) for our observing its existence. Here, however, the appendix isn't causally efficacious *qua* appendix. Perhaps, if pressed, this example becomes problematic for making my point. The point, however, is that in principle there could be effects of physical causes that are not causally efficacious (perhaps we would never know about them). Such effects would fail to be physical on this account. If one modifies the account to include the effects of all physical causes, this would obviously be too permissive. In fact, Hasker believes that immaterial substances are caused to exist from complex physical causes.

1.1.3 Publicly Observable Criterion

Another recent attempt to characterize physicalism uses the standard of what is *publicly observable*. What is publicly observable is typically characterized by truths that can be known from multiple perspectives. Often, this is considered to be a criterion for objectivity. Daniel Dennett suggests this account of physicalism when he writes, "I declare my starting point to be the objective materialistic, third-person world of the physical sciences."[24] Richard Swinburne has a similar understanding of physicalism as events that are publicly observable and do not require privileged access.[25] (Swinburne defines events that can be known by "privileged access" as those where only one person is able to know about these events directly by experiencing those events.)

Defining physicalism in terms of what is publicly observable suffers from a number of problems. First, one must be careful how to define what is publicly observable. If observability is restricted to observations that can be made without instruments, then this account fails criteria (i) since electrons, quarks, and other subatomic entities cannot be observed, although they are paradigmatic physical things. On the other hand, if observability is broadened to include entities that can be observed indirectly (such as through the effects of directly unobservable entities), it will fail criteria (ii) since there is no good reason to

---

[24] Dennett (1987), p. 5.

[25] Swinburne (1996), p. 71.

suppose Cartesian souls, acts of God, and irreducible qualitative conscious states cannot be observed indirectly, although they are clearly not physical things. So, the observability criterion has a serious problem with its primary standard for demarcating the physical from the non-physical.

But even granting that there is a solution to the problem of defining what is publicly observable such that it can include indirectly observable subatomic entities and not include acts of God and Cartesian souls, this criterion still faces severe problems. First, this view cannot adequately meet the second criterion. This account of physicalism provides objectionable results in a world with telepathic beings, like the Betazoids, portrayed most notably by Deanna Troi in *Star Trek: The Next Generation*.[26] The Betazoids have telepathic and empathic powers that allow them to observe directly the conscious states of other beings. This would mean that many of the phenomena we typically consider to be essentially private would be classified as physical in a world with creatures like the Betazoids. Perhaps, the fix to this problem is to argue that "publicly observable" should not be interpreted as "whatever is logically possible to be observed by others," but rather according to what could be observed by others according to nomological or metaphysical possibility. This, however, admits of the difficulty of determining what exactly should count as nomologically or metaphysically possible to observe publicly. At the very least, it isn't clear that

[26] Technically, Deanna is only half Betazoid, but full-blooded Betazoids, such as Deanna's mother Lwaxana Troi, appear in several episodes. Similar concerns about the observability criterion are raised in Maddell (1988), p. 4.

Betazoids are ruled out by those modalities. (For all we know, we will discover the existence of Betazoids in our world 500 years from now.) A second problem that remains even if an acceptable modal interpretation can be determined is that defining physicalism by what is publicly observable does not satisfy criterion (iii). By picking out physical things as publicly observable, it does no work to explain the structural dependence and relations of physical things. So, I will not take the publicly observable account of physicalism to be adequate for my purposes.

It is tempting to modify the public observability criterion to state the physical in terms of what is not observable by the method of introspection. But this modification is of no avail. No physicalist would accept this account because it blatantly presumes that there are some things that exist which are not physical (since we know some things exist through introspection). So, this definition would beg the question against physicalism. Furthermore, it isn't clear that it would satisfy criterion (ii), since God would presumably qualify as a being that isn't observable by introspection, and which is clearly not physical. Also, like the original version of the publicly observable criterion, it fails the third criterion since there are no causal or explanatory relation between microphysical entities and macrophysical objects.

1.1.4 Supervenience Criterion

Another recent attempt to define the concept of physicalism uses the notion of supervenience.[27] On this account, physicalism is true if everything that exists supervenes on the physical. Supervenience, stated crudely, is a relation between two sets of properties, base properties and supervenient properties, where the base properties determine the supervenient properties. The basic idea is that there can be no difference in supervenient properties without a difference in base properties. Put in a slightly different way, if supervenience holds, then no two objects with the same base properties could differ with respect to their supervenient properties. Although supervenience was introduced to twentieth century philosophy by G. E. Moore as a relation to characterize a non-natural account of moral realism,[28] ever since Donald Davidson has used the term in the philosophy of mind it has become commonly used to define contemporary physicalism.[29]

Physicalists have employed supervenience in various different ways to define physicalism. As a token example of the way many physicalists use supervenience to understand the concept of physicalism, consider John Haugeland's weak supervenience principle, "The world could not have been different in any respect, without having been different in some strictly physical

---

[27] See the essays in Kim (1993a) and McLaughlin and Bennett (2008) for an overview of the current philosophical views on supervenience.

[28] Moore (1903).

[29] Davidson (1970).

respect—that is, in some respect describable in a canonical language of physics."[30] Haugeland expresses his supervenience principle more precisely as

> K *weakly supervenes on* L (relative to W) just in case any two worlds in W discernible with K are discernible with L.[31]

Accounts of physicalism that employ something like Haugeland's weak supervenience are unsatisfactory because they do not adequately satisfy criteria (ii) and (iii). While weak supervenience might describe a correlation between the base property and the supervenient property, weak supervenience has no guarantee that all the *relata* in the supervenience relation will be physical in nature. More bluntly, a definition of physicalism based on weak supervenience does not rule out the possibility that Cartesian souls, irreducible qualitative mental states, or even God could supervene on the base conditions. After all, many dualists believe that there is a law-like covarying relation that holds between physical brain states and irreducible mental states, which would fit the concept of physicalism given in terms of weak supervenience.

Weak supervenience physicalism also fails to satisfy the third criteria for an account of physicalism. Weak supervenience provides no explanation of the structural and causal dependence that holds between the base and supervenient properties. Even if weak supervenience physicalism satisfied the first two criteria, it would not provide any explanation of the connection between the base

---

[30] Haugeland (1984), p. 1.

[31] Haugeland (1982), p. 97.

and supervenient properties. This problem has been illustrated through Jaegwon Kim's well-known ammonia molecule problem.[32] Weak supervenience is compatible with the possibility that there could be two worlds that are identical in all microphysical respects except that one world has a minor difference (such as having one extra ammonia molecule around one of the rings of Saturn), but they could differ radically in the supervenient properties (e.g., one world could have human beings with conscious experience, whereas the other world could have no conscious experience at all). The reason weak supervenience is exposed to this problem is because weak supervenience offers no explanation as to how the base properties stand in the supervenience relation with their respective supervenient properties.

Perhaps, one maybe tempted to adjust and qualify the account of supervenience to fix the aforementioned problems. I am inclined to resist this maneuver since similar problems are bound to be lurking for any account of physicalism that essentially relies on supervenience. My reason for this general worry about the adequacy of using supervenience to define physicalism is due to the *kind* of relation that supervenience is. Supervenience on its own does not characterize the explanatory, causal, or dependence relations between its *relata*. For this reason, Kim claims that supervenience is a statement of the mind-body problem, not a solution to it:

---

[32] Kim (1993b).

> Supervenience itself is not a *type* of dependence relation—it is not a relation that can be placed alongside causal dependence, mereological dependence, dependence grounded in definability or entailment, and the like. It is not a metaphysically deep, explanatory relation, being only a phenomenological relation about patterns of property covariation. If this is right, mind-body supervenience *states* the mind-body problem—it is not a solution to it. Any putative solution to the problem must, at a minimum, specify a dependence relation that grounds mind-body supervenience. We expect mind-body theories to be explanatory theories.[33]

Thomas Nagel expresses a similar objection to employing supervenience alone to characterize physicalism:

> We have good grounds for believing that the mental supervenes on the physical – i.e. that there is no mental difference without a physical difference. But pure, unexplained supervenience is not a solution but a sign that there is something fundamental we don't know. We cannot regard pure supervenience as the end of the story because that would require the physical to necessitate the mental without there being any answer to the question *how* it does so. But there *must* be a "how," and our task is to understand it. An obviously systematic connection that remains unintelligible to us calls out for a theory.[34]

One might still try to make supervenience work by detailing the explanatory or dependence relation in a specified account of the supervenience relation. For example, one might define the physicalist's notion of supervenience as "*A* physically supervenes on *B* when any two lawfully identical worlds with *A* are indiscernible with respect to *B* whenever *A* satisfies dependence relation *R* with respect to *B*." One problem with this sort of approach is that supervenience is no longer doing any work in the account. One might as well simply stick to

---

[33] Kim (1997), p. 190.

[34] Nagel (1998), pp. 344-345.

specifying the concept of physicalism in terms of the dependence relation since that relation is the one that is providing the explanatory connection from the base properties to their supervenient ones.

To see this point, recall G. E. Moore's non-natural moral realism. One could define Moore's moral theory using supervenience: all moral properties (of a certain specification) supervene on physical properties (of a certain specification). But Moore clearly did not think that moral properties are physical properties because they stand in this supervenience relation to each other. As we know, Moore believed that his open question argument established that moral properties cannot be identical with physical properties. After all, if moral properties were identical with certain physical properties, it would be superfluous to use supervenience to state the theory. The point behind stating one's theory using supervenience is to describe a covariation in different *kinds* of properties. Many so-called non-reductive "physicalists" are content to state physicalism in terms of supervenience alone. However, such a view would permit Moore's non-natural moral realism to count as being physical. Supervenience by itself, then, cannot provide a plausible account of physicalism.[35]

Although it is very likely that the right account of physicalism will imply some sort of supervenience thesis, supervenience in itself does not provide a satisfactory account of physicalism. In fact, most attempts to define physicalism

[35] See Bealer and Koons (2010), p. xvi for a similar criticism of brute supervenience physicalism.

by using supervenience alone are compatible with property and substance dualism. However, even if supervenience is not sufficient by itself to characterize physicalism, it is most likely a necessary condition for physicalism. Given that more is needed to give an adequate account of physicalism, for my purposes it will not suffice to characterize physicalism merely as a supervenience claim. An adequate account will also provide an explanation for why supervenience holds between the microphysical bases and the truths that supervene on them.

1.1. 5 Constitution Criterion

One candidate relation for providing an explanatory connection between the base properties and supervenient properties is the relation of constitution, that is the relation that holds between constituent parts and their wholes. There is a sense in which wholes have their properties in virtue of their parts. Additionally, wholes stand in some sort of causal dependence to their parts. For example, a bridge made of wooden planks exists in virtue of its wooden planks (take away the wooden planks, you take away the bridge). Furthermore, the bridge's properties, such as being a specified length, being able to support a certain weight, etc., all are explicable in virtue of the relations that hold between the properties of the wooden planks. The bridge is different from the wooden planks; the bridge's properties aren't the exact same properties as the properties of any individual wooden plank, but the bridge's existence and its properties are

explained by the properties (and relations) of its parts. So, it seems that constitution is a prime candidate for the sort of relation that underwrites criterion (iii), the explanatory bridge between micro-entities and macro-entities for an account of physicalism.

One problem with constitution is that it is not enough to rule out novel emergent properties that would remain unexplained by a complete account of physics, thereby violating criteria (ii) and (iii). For example, many who adopt a material constitution view of human persons often embrace the result that many properties of the whole that supervene on the aggregate parts are not explicable by the parts.[36] Typically, the properties associated with consciousness are taken to be novel, irreducible properties of a person (the whole) that exist in virtue of being constituted by a person's body (the parts). But unlike the macroproperties of the bridge, the causal powers of these conscious properties of persons are typically not taken to be explicable in terms of the properties and relations of the parts. Rather, these new properties are intrinsic properties of the whole—they are not reducible to the relational properties of their parts. But constitution doesn't explain the novel, irreducible, intrinsic macroproperties of the whole.[37] Therefore, defining physicalism in terms of constitution does not explain the

[36] Such as Baker (2000), Pereboom (2002), and Merricks (2003), especially pp. 85-117.

[37] For more on the explanatorily thin nature of the constitution relation as it applies to the mind-body problem, see Kim (1998), pp. 18-19.

relation between properties of wholes and parts in a way that is guaranteed to be consistent with physicalism.

So, even if physicalism implies that persons are constituted by their physical parts as a necessary condition for physicalism, it is not sufficient for physicalists to appeal to the constitution relation that holds between mind and body to *explain* the mental properties. As we've seen, the constitution relation is compatible with the emergence of intrinsic, irreducible properties that apply to wholes in virtue of being constituted by their parts. But these intrinsic properties of the wholes are not necessarily explained by the properties and relations of the aggregate parts.

1.1.6 Causal/Functional Realization Criterion

Causal realization is the primary way that functionalist theories of mind characterize mental states. Mental states are not defined by their intrinsic character, claim functionalists, but rather mental states are what they are in virtue of the functional role they play in the system of which they are apart. For example, a crude functionalist account of pain would define pain as the sort of process in a system that is typically caused by damage to the system, indicates that damage has occurred in the system, causes the system to desire to be out of that state, and causes wincing or moaning from the system. The key insight for providing an account of mental states by functional realization is that the mental

states are characterized by the causal role they play in the system, not by their intrinsic features.

Functionalism is not necessarily committed to a physicalist account of the world. However, many physicalists have found functional realization to be a fruitful tool in developing a thoroughgoing physicalism. If all of the realizers of a functional system are physical, then the functionally realized state is physical since there is nothing more to being in that state than having those causes. In other words, there is nothing "over and above" being a functionally realized state than having the physical properties that causally realize being in that state.[38] Of course, there are many different accounts of physical functionalism, but I have in mind the sort generally associated with David Armstrong,[39] David Lewis,[40] Frank Jackson,[41] Jaegwon Kim,[42] David Chalmers,[43] and Andrew Melnyk.[44] (I would like to add that I do *not* have in mind the sort of functionalism employed by non-reductive physicalists, such as Sydney

---

[38] The "over and above" locution echoes the phrase used by Smart (1959) to suggest that brain states are identical to mental states, rather than merely correlated with each other. In the same vein, I am suggesting that functionally realized states are nothing more than having the states that causally realize them.

[39] Armstrong (1968).

[40] Lewis (1966), (1992), (1994).

[41] Jackson (1998).

[42] Kim (1998), (2005).

[43] Chalmers (1996). Chalmers is a property dualist, but he does characterize physicalism as a reductive *a priori* thesis.

[44] Melnyk (2003), although Melnyk is not an *a priori* physicalist.

Shoemaker.[45] My reasons for doing so will be apparent when I discuss the causal exclusion argument later.) Since there are bound to be differences among these advocates of functional realization, I will use Kim's view as a token example since I find his approach to be one of the most clear and promising accounts.

Kim defines a functional property as a second-order property that is realized when the right sort of first-order property is in the right conditions.[46] The property of being in pain, for example, is realized by a system with the right first-order property (e.g., the right sort of nervous system) when the right conditions obtain (e.g., damage occurs to the system) where these conditions specify that the first-order property has pain's typical causes and effects. In order to see how functional properties provide an explanatory link in an account of physicalism, it is important to understand more clearly Kim's notion of a second-order property. Kim offers the following analysis of second-order properties:

> *F* is a *second-order property* over set **B** of base (or first-order) properties iff *F* is the property of having some property *P* in **B** such that *D*(*P*), where *D* specifies a condition on members of **B**.[47]

The base or first-order properties do not necessarily have to be first-order in any absolute sense. The idea in this analysis is that the base properties are first-order relative to the second-order properties that they realize (in conjunction

---

[45] Most recently in Shoemaker (2007). See Churchill and O'Connor (2010), for specifics on how Shoemaker's position is susceptible to the causal exclusion argument.

[46] Kim (1998), pp. 19-23.

[47] Kim (1998), p. 20.

with condition *D*). Kim restricts condition *D* to causal/nomic relations. After all, the heart and soul of functional realization is accounting for certain higher-order properties in terms of the causal roles they play in a system, and Kim's condition *D* provides this aspect to his account of functional realization.

Kim's account of physical realization will entail supervenience between brain states and mental states.[48] If *P* realizes the second-order property *M* in system *S*, it follows from Kim's account of functional realization that *P* nomologically necessitates *M* in *S*. For various token instances of *P*, that is <P_1, P_2, . . . , P_n>, each will realize a token instance of *M*, that is <M_1, M_2, . . . , M_n> in *S*. Consequently, the *M*s nomologically supervene on the *P*s. The upshot of all of this for my current purposes is that functional realization will supply an explanation as to why the mental properties supervene on the physical properties, thereby satisfying criterion (iii). Furthermore, this position can satisfy criteria (i) and (ii) for an account of physicalism. The account can include all of the standard physical things that ought to be in a physicalist ontology, while ruling out all of the paradigmatic non-physical things. Thus, physicalism can be understood in the following way:

> Physicalism is true iff everything in the world is either (i) a fundamental constituent in physics, or (ii) supervenes on the fundamental constituents of physics by being functionally realized by base properties that are either fundamental constituents of physics or are properties that are eventually realized by fundamental constituents of physics.

[48] Kim (1998), pp. 23-27.

I believe that this provides a sufficient sketch of the basic ingredients that I take to satisfy my three criteria for a plausible account of physicalism. The positive ontological adequacy criterion is satisfied because the conjunction of the fundamental physical constituents and physical realization is sufficient to include all the paradigm physical things either as fundamental constituents or as things realized ultimately by those constituents. The negative ontological adequacy criterion is satisfied because the account of physical realization that I take to be the most plausible view of physicalism does not allow novel, irreducible properties to emerge or supervene on systems. This restriction follows because the account of functional realization characterizes second-order properties in terms of physical causal laws and relations of the first-order properties. Importantly, causal realization, as I am using the term, restricts higher-order properties from possessing intrinsic features that are causally and relationally independent of the lower-level properties and relations. Finally, the explanatory adequacy criterion is satisfied because the relation of functional realization is an explanatorily deep relation. On my account of physicalism when *X* is functionally realized by *Y*, we have an explanation why *X* has the features that it does—*X* is what it is by virtue of the physical causal laws and relations that apply to the physical and relational properties *Y* possesses. In fact, on my understanding of physicalism, having *X* is explained by having *Y* because having *X* is nothing more than having *Y*. As the saying goes, having *X* is nothing "over and above" having *Y*.

Thus far, I have maintained that physicalism, in order to be a plausible view, must satisfy three criteria: (i) positive ontological adequacy, (ii) negative ontological adequacy; and (iii) explanatory adequacy. After considering various candidate ways for meeting these criteria, I have settled on a specific account of functional realization that satisfies these criteria. Below, I will expound on two consequences of this approach that are relevant to my project, and then I will provide a brief argument to further motivate this analysis of physicalism.

1.1.7 Reductive and A Priori Physicalism

Two consequences of my account of physicalism are that physicalism has a reductive and *a priori* nature. As to what I mean by saying physicalism has these features and why this is so shall be the subject of this section.

Reductionism is a slippery concept. In its most general form, reductionism refers to analyses that account for one thing in terms of another more basic sort of thing. Strict forms of mind-body reduction, for example, have tried to explain mental properties by appealing to bridge laws that are construed as biconditional statements such as being in pain occurs if and only if brain state *P* occurs.[49] This sort of reduction is widely regarded to be untenable due to the problem of multiple realization.[50] If being in pain (for example) is multiply

[49] See, for example, Nagel (1961), ch. 11.

[50] See Fodor (1974).

realizable by different base properties, then biconditional bridge laws will fail to account for being in pain.

The sort of reductionism that I have in mind does not depend on biconditional bridge laws to provide a reductive account of mental states in terms of physical states. Rather, to reduce something on my view is to provide an account of how it is functionally realized.[51] In other words, being in pain is reduced to instantiations of its functional types. If being in pain is realized by physical states $<P_1, P_2, \ldots, P_n>$, then there is nothing more or less to being in pain than being in state P_1, or P_2, or . . ., or P_n. The causal powers of being in pain are nothing more or less than having the causal powers conferred by being in state P_1, or P_2, or . . ., or P_n. One consequent of this account of functional reductionism is the reductions between mental states and their realizers will not be strict identities.[52] Strict identities will not hold for functionally reduced *relata* because functional reductions hold true relative to the structures of systems that implement the base properties and to the laws of nature in a given world.[53]

Furthermore, providing a reductive account of physicalism in terms of functional realizability does not have the problem of accounting for the possibility of mental states being multiply realizable. On the functionalist approach to reduction, being in pain (for example) is to be in a state meeting

---

[51] Following Kim (1992a); (1998), pp. 23-27; (1999); (2005), pp. 22-29.

[52] In other words, they will not be rigid designators in Kripke's (1980) sense.

[53] The upshot is that functional reductions provide semi-rigid designators, that is, they remain referentially successful across nomologically fixed worlds.

certain causal specifications we can call *C*. If many different physical states <P_1, P_2, . . . , P_n> meet the specifications of *C*, then being in pain can be realized by many different states in virtue of these states having the same causal powers relative to a system and a set of physical laws. This is the sense in which I claim physicalism is a reductive thesis.

A second feature of the account of physicalism I take to be most plausible is that physicalism turns out to have an *a priori* thesis. Frank Jackson is best known for representing *a priori* physicalism as the claim that the complete set of true propositions about the fundamental physics of a world entails *a priori* the complete set of true propositions about the world.[54] In other words, if *W** stands for the complete set of true propositions about microphysics and *W* stands for the complete set of true propositions about the world, then "if *W**, then *W*" will be an *a priori* entailment. Of course, many true propositions in *W** and *W* will only be knowable *a posteriori* through empirical methods. However, the conditional "if *W**, then *W*" will be an *a priori* entailment. This follows from seeing that if physicalism is true, then the higher-order referents in propositions that constitute *W* will refer to things that can ultimately be reduced by appealing to functionally realizable processes that are constituted by entities in *W**.

---

[54] Jackson (1995), (1998), and (2007a). See also Chalmers and Jackson (2001). Some arguments for *a priori* physicalism will be discussed under the section "The Physical Knowledge Intuition" below.

1.1.8 The Causal Exclusion Argument

Undoubtedly, many physicalists will be uncomfortable with my characterization of their position. They might insist that I am demanding too much for physicalism to be true or that it is a caricature. To the contrary, I would point out that I have given three criteria that motivate and support my understanding of physicalism. Furthermore, my portrayal of physicalism is more-or-less understood along the same lines of important physicalists like David Armstrong, David Lewis, Frank Jackson, Jaegwon Kim, and Andrew Melnyk, for example. In addition to these justifications for my account of physicalism, I will offer the causal exclusion argument as another reason to accept my account of physicalism. Since the causal exclusion argument employs premises that are widely amenable to physicalist sympathies, and it implies that physicalism should be understood along the lines I have specified above, I will take this as additional and independent grounds for accepting my account of physicalism.

The causal exclusion argument has been pressed most fervently by Jaegwon Kim against non-reductive accounts of physicalism.[55] The appeal of the argument stems from a few basic principles that all physicalists should find

[55] Kim (1992b), (1998), and (2005).

appealing.[56] The first principle is the causal closure of physical domain, which Kim defines as follows:

> *The causal closure of the physical domain.* If a physical event has a cause at *t*, then it has a physical cause at *t*.[57]

The basic idea behind this principle is that all physical events can be given a sufficient causal explanation in terms of physics. In other words, there is no need to look for or hypothesize a non-physical cause for any putative physical event. Denying the causal closure of the physical seems tantamount for denying physicalism. To allow violation of the causal closure principle is unacceptable to most physicalists because doing so concedes that a complete and comprehensive physical account of all physical phenomena cannot be accomplished, and this is something that Kim claims "no serious physicalist could accept."[58] Some, like Karl Popper, have used the causal closure of the physical as the primary way to define physicalism.[59] For these reasons, the causal closure of the physical stands as a principle that most physicalists are likely to accept.

Kim's next principle is the principle of causal exclusion.

---

[56] Additionally, the argument presumes that causation is robust in such a way that deflationary accounts of causation are untenable. See Churchill and O'Connor (2010) for a brief defense of this account of causation in the context of the causal exclusion argument.

[57] Kim (2005), p. 15.

[58] Kim (1998), p. 40.

[59] Popper and Eccles (1977), p. 51: "the physical principle of the closedness of the physical . . . is of decisive importance, and I take it as the characteristic principle of physicalism or materialism."

> *Causal exclusion principle*. If an event *e* has a sufficient cause *c* at *t*, no event at *t* distinct from *c* can be a cause of *e* (unless this is a genuine case of causal overdetermination).[60]

Kim generalizes the causal exclusion principle to apply more generally to causes of generation and determination:

> *Principle of determinative/generative exclusion*. If the occurrence of an event *e*, or an instantiation of a property P, is determined/generated by an event *c*—causally or otherwise—then *e*'s occurrence is not determined/ generated by any event wholly distinct from or independent of *c*—unless this is a genuine case of overdetermination.[61]

The basic motivation behind the exclusion principles comes from recognizing that there is strong causal and explanatory dependence that holds between certain causes and events, especially causes that determine or generate higher-order properties (such as the dependence relation defined earlier as functional realization).

The combination of causal closure of the physical domain and the causal exclusion principles can be used to wreak havoc on allegedly emergentist and non-reductive accounts of physicalism. The exclusion argument, as it is commonly called, begins by taking an alleged token of mental causation, where one mental event *M*, putatively causes another mental event *M**.

(1) Mental event *M* causes mental event *M**.

This claim seems plausible. It is widely believed that one mental state can cause another. For example, my thought of a piece of cheesecake can cause me to have

---

[60] Kim (2005), p. 17.

[61] Kim (2005), p. 17.

a desire for a piece of cheesecake. Or, perhaps, one might think that as one considers the content of one's thoughts that make up the premises of a valid argument, there is a sense in which the contents of those thoughts are considered to be the cause of one's thought that the conclusion follows from those premises.

Another reason to think that mental causation occurs is that if mental states cannot cause other mental states, then there seems to be no basis for human agency, or any hope for acquiring knowledge.[62] For human agency, we suppose that our mental states make a causal difference in controlling our physical behavior. In order for us to have knowledge of the external world, it is necessary for our perceptual mental states, the beliefs we have about them, and the inferences we make from them to be causally efficacious. Therefore, if our mental states are causally impotent, then it seems we have no basis for accepting that we are in any sense in control of our behavior or possessors of knowledge. So, denying the efficacy of mental causation carries a hefty toll.

Given a commitment to some sort of physical-mental supervenience, some form of which all plausible forms of physicalism either explicitly endorse or tacitly imply, the following two causal claims must be accepted by physicalists.

(2) Physical event *P* causes *M*.

(3) Physical event *P** causes *M**.

---

[62] These consequences of denying mental causation have been drawn from Kim (2005), pp. 9-13.

In this case, *P* and *P** refer to the brain states that causally realize the mental states *M* and *M** respectively. Claims (2) and (3) should be understood also to provide a fully causal and generative explanation of the existence and determinable properties of *M* and *M**. In other words, *P* alone is sufficient for *M*, and *P** alone is sufficient for *M**.

Furthermore, in the case of supposedly non-reductive or emergent physicalisms, mental states are not identical with the functional states that realize them, so the following non-identities will hold:

(4) $M \neq P$

and

(5) $M^* \neq P^*$.

With the non-identities implied by non-reductive physicalism made explicit, now we can see that there are two distinct causes for *M**. In (1) *M* is the alleged cause of *M**, and in (3) *P** is the alleged cause of *M**. The exclusion principles noted above, however, do not allow multiple causes or explanations to account for one event (unless it is genuinely a case of overdetermination, which this case isn't—this will be addressed below). Perhaps, one might think a solution can be found by the following maneuver:

(6) *M* causes *M** by causing its supervenience base *P**.

In this way, one might think that this avoids the problem of causal overdetermination. However, this pushes the causal overdetermination back a step. According to the causal closure principle, *P** must have a physical cause.

For the sake of keeping this example as simple as possible, let's allow that the physical cause of *P** is *P*, which is the next claim in the causal exclusion argument.

(7) *P* is the cause of *P**.

But now we have causal overdetermination of *P**. According to (6) *M* is the cause of *P**, and according to (7) *P* is the cause of *P**. Since *P* and *M* are distinct entities on non-reductive physicalism, the non-reductive physicalist cannot claim that *M* causes *M** by *P*'s causing *P**. Given causal closure, (7) is non-negotiable for the physicalist. Given non-reductive physicalism's commitment to (4) and (5) and the exclusion principles, the non-reductive physicalist is forced to deny (6). The picture that the non-reductive physicalist is forced to accept, then, is given by the conjunction of (2), (3), (4), (5), and (7). Claims (2) and (3) describe the supervenience of the mental on the physical. Premises (4) and (5) are essentially what make a non-reductive account of physicalism to be non-reductive. Statement (7) affirms the causal relation between *P* and *P**. The consequence is that *M* and *M** are causally ineffective. Given the importance of and evidence for mental causation, this significantly weakens the plausibility of non-reductive physicalism. Of course, reductive physicalism does not have this problem since it allows for *P*=*M* and *P**=*M**; *M* is causally efficacious because *P* is causally efficacious.

Perhaps the non-reductive physicalist will try to resist the causal exclusion argument by allowing for the causal overdetermination of *P** by both *P* and *M*.

Kim's response is to emphasize that in genuine cases of causal overdetermination each overdetermining cause plays a distinct and distinctive causal role.[63] Typically, this can be seen by each overdetermining cause taking separate, independent causal chains that happen to coincide on a common effect. However, since all physicalists take supervenience to hold between *P* and *M*, they do not satisfy the coincidental story that typifies overdetermination. If *M* is not reducible to *P* in terms of functional realization, in what way is *M* causally contributing to *P** that is not already being contributed by *P*? The problem isn't merely that there would be overdetermination, but rather the problem is that any putative causal efficacy we would want to attribute to *M* has already been accounted for by the causal efficacy *P*. Since the only causal efficacy that *M* could have on *P** is by the exact same causal process conferred by *P*, this makes the overdetermination option especially unpalatable.

So, I have provided an analysis of physicalism where it amounts to claiming that everything in the world is either a fundamental constituent of physics or ultimately realized by fundamental constituents in physics. I have supported a reductive analysis of physicalism by appealing to three criteria that any account of physicalism should meet, showing that important physicalists have held similar views, and arguing from the causal exclusion argument that non-reductive physicalism cannot account for the causal efficacy of mental states.

---

[63] Kim (2005), pp. 46-52. Here the rejection of deflationary accounts of causation is crucial.

Next, I will draw an implication from this analysis, which I believe establishes a condition whereby physicalism can be falsified.

1.1.9 The Physical Knowledge Intuition

As we shall see more clearly in the next chapter, there is a crucial intuition that underwrites the Knowledge Argument for dualism. The basic idea is that if physicalism is true, then knowing all the physical truths about the world is sufficient for knowing all the truths about the world. This seems to be a natural consequence of the reductive and *a priori* aspects to physicalism that I have labored to establish in the prior sections of this chapter. Given the importance of this intuition for the Knowledge Argument, I will use this section to elaborate on this intuition and its support.

As I have characterized it, if physicalism is true, then everything that exists is either a fundamental constituent of physics or reducible to fundamental constituents of physics by being causally realized by them. The upshot of this analysis is that there is nothing "over and above" the fundamental constituents of physics. So, "being a rock" is nothing over and above the functional state that causally realizes "being a rock." Likewise, physicalism implies that "being in pain" is nothing over and above the functional state that causally realizes "being in pain." All of this lends credence to the notion that if physicalism is true, then knowing all the fundamental physical truths about the world is sufficient for knowing all the truths about the world. Once again, knowing that an object is a

rock is satisfied by knowing that certain fundamental constituents of physics are causally related to one another in the right way (since being a rock is nothing more than physical parts being in certain functional state). Similarly, knowing that fundamental constituents in physics are arranged in the right sort of causal system should be sufficient for knowing that one is in pain (since being in pain is nothing over and above the causally realized system).

These insights support the idea that if physicalism is true, then knowing all the truths about the fundamental physical constituents of the world is sufficient for knowing everything about the world. The contrapositive of this conditional claim, however, generates a sufficient condition for concluding that physicalism is not true. I will call this the Physical Knowledge Intuition:

> (PKI) If complete possession of all knowledge of physical truths isn't sufficient to provide all propositional knowledge of the actual world, then physicalism is false.[64]

There are a few points I need to clarify regarding the PKI. In the antecedent, I use the locution "propositional knowledge," which I mean to be a kind of knowledge that fits the schema *S* knows that *p*. The idea is that propositional knowledge involves the application of concepts and where the content of one's knowledge can take a truth-value (which in genuine cases of knowledge must be a true truth-value). Propositional knowledge can be contrasted with know-how. Know-how refers to one's knowledge to exercise certain abilities, such as an

[64] A caveat: one kind of propositional truth that should not be included among those knowable from the complete set of physical truths are truths that essentially involve indexical content. See §5.1 for more details.

accomplished golfer's knowledge of how to drive a golf ball over 300 yards.[65] Presumably, it is possible to have know-how without propositional knowledge of each step needed to perform an ability and vice versa.

Since the PKI only requires one to be able to deduce all propositional knowledge of the world from all the physical truths, it allows physicalists to concede quite reasonably that knowing all the physical truths about the world will not deductively entail all knowledge about the world. For it doesn't seem correct that all knowledge of the world (including know-how) should follow from the complete set of physical truths. Whether this concession is sufficient to save physicalism from the perils of the Knowledge Argument will be considered and refuted in chapter four.

In addition to the aforementioned reasons for accepting the PKI on the grounds that it follows from my analysis of physicalism, there are some independent reasons for thinking that the PKI is true. For example, Richard Swinburne argues that if one knew every event that took place in the world, then one would know all that happened in the world.[66] Swinburne claims that an event is a particular substance at a particular time possessing a particular property or relation. If physicalism were true, though, all the events would

---

[65] In order to satisfy the schema "*S* knows how to [*verb*]", it does not necessarily require the subject to have conceptual content about one's ability to [*verb*]. A subject may satisfy the conditions for possessing know-how by virtue of having certain abilities even if the subject has no mental states at all.

[66] Swinburne (1996), p. 70-73. Swinburne speaks of monism, rather than physicalism. I have modified his terminology to fit mine.

consist in substances and properties that are most fundamentally characterized by physics. Therefore, Swinburne concludes, that if physicalism were true, then knowing every physical event that takes place in the world would be sufficient for knowing everything about the world. Swinburne's line of reasoning can be reconstructed in the following way:

(8) If one knew all the events that had happened, then one would know all that had happened.

(9) If physicalism is true, then all the events that occur in the world are most fundamentally physical.

Therefore,

(10) If physicalism is true, then knowing all the physical events would be sufficient to know all that had happened.

Since the contrapositive of (10) is almost virtually identical to (PKI), the reasonability of the PKI is supported by the plausibility of (8) and (9).

Perhaps, critics may allege that the PKI is too strong. Even so, it is possible to formulate a restricted version of the claim for the purposes of the Knowledge Argument. Following Nagasawa and Stoljar, we can call this restricted version of the PKI the *psychophysical conditional*.[67] The psychophysical conditional states that all the physical truths of the world imply all the psychological truths of the world. If any form of physicalism implies that the psychological features of our world supervene on the physical features of our world, then the psychophysical condition will be taken to be a necessary truth by

[67] Nagasawa and Stoljar (2004), pp. 14-16.

those physicalists. While it isn't controversial that physicalism entails that the psychophysical conditional is a necessary claim (because it is uncontroversial that physicalism entails psychophysical supervenience), the more controversial claim is that the psychophysical condition is an *a priori* necessary truth. Since the so-called new theories of reference in the philosophy of language,[68] it has become fashionable for physicalists to take the psychophysical condition as an *a posteriori* necessary truth. I will offer some brief reasons to motivate the position that physicalists should take the psychophysical conditional to be an *a priori* necessary truth.

The reasons for physicalists to take the PKI or the psychophysical conditional to be a necessary *a priori* claim are directly related to the reasons that physicalism has the *a priori* component that I have discussed previously. Frank Jackson is the most prominent defender of *a priori* physicalism, so I will rely on his work to support the claim that the psychophysical conditional is an *a priori* necessary claim.[69] Even working with the so-called new theories of meaning and reference that are currently fashionable in philosophy of language, Jackson argues that physicalists should take the psychophysical conditional to be an *a priori* necessary truth.[70] As a first step, he introduces the following claim:

---

[68] Primarily due to the work of Putnam (1975b) and Kripke (1980).

[69] See especially Jackson (1995), (1998), and (2007a).

[70] I believe that the new theories of meaning and reference are mistaken. My reasons for concluding this will be laid out in §5.2. In the meantime, I will try to motivate *a priori* physicalism without presuming that these theories are false.

(11) H_2O covers most of the planet.

From (11), can we conclude that water covers most of the planet? In one sense, Jackson grants, we could validly deduce from (11) that water covers most of the planet because the conditional claim, "if H_2O covers most of the planet, then water covers most of the planet" is necessarily true, although it isn't knowable or deducible in any *a priori* way. So the inference from (11) to the claim that water covers most of the planet is valid in the sense that it is necessarily truth-preserving. However, it is invalid in the sense that it is not possible for one to deduce *a priori* from (11) the conclusion that water covers most of the planet. So, (11) strictly implies the conclusion that water covers most of the planet, but it does not *a priori* entail it.

Jackson, next, considers a case where the information we have about water and H_2O is more robust. In addition to (11), consider how the argument would go with

(12) H_2O fills the water role.

Now (11) and (12) together entail *a priori* that water covers most of the planet. When the information about H_2O is rich enough—that is, with both (11) and (12)—the knowledge that water covers most of the planet can be considered valid in both senses described above. It is valid in the first sense of being necessarily truth-preserving, and it is also valid in the sense of being deducible *a priori*.

Like the case of being able to deduce *a priori* that water covers most the

planet from (11) and (12), Jackson thinks that one can deduce *a priori* truths of human psychology from a sufficiently rich account of physical truths about human neurophysiology. What Jackson has in mind is something like the following:

(13) Mental state *M* = the state that plays the causal role *R*.

(14) The state that plays the causal role *R* = brain state *B*.

Together, (13) and (14) are sufficiently rich enough to provide an *a priori* deducible inference to the claim that *M* = *B*. If mental states can be reduced using functional realizations (as I have urged physicalists to do above), then physicalists should be prepared to accept the *a priori* inference from (13) and (14) to the conclusion that *M* = *B*. Physicalists who resist this inference do so because they are skeptical whether it is possible to reduce mental states in terms of causal realization—a position which I have already argued has significant problems and that I think is best to be avoided. So, given the commitments I have already specified for the physicalist, it is reasonable to think that the psychophysical condition and the PKI are both plausible claims that either will stand as an important part of explaining why physicalism is true or will demonstrate why physicalism is untenable.

A final reason to think that physicalism should accept the conditional claims in the PKI and the psychophysical condition as *a priori* necessary truths is inspired by Jackson's observation about simple organisms:

It is implausible that there are facts about very simple organisms that

> cannot be deduced a priori from enough information about their physical nature and how they interact with their environments, physically described. The physical story about amoeba and their interactions with their environments is the whole story about amoeba. . . . But according to materialism, we differ from amoeba essentially only in complexity of ingredients and their arrangements. It is hard to see how that kind of difference could generate important facts about us that in principle defy our powers of deduction . . . .[71]

This brief argument provides a reason for physicalists to accept something like the PKI as an *a priori* necessary truth for reasons that are independent of the motivations that I have included in my account of physicalism. For example, even a non-reductive physicalist or one who is inclined to accept a constitutional account of physicalism should feel the lure of Jackson's brief argument. If human beings differ only in degree from very simple organisms and all the truths about very simple organisms can be deduced *a priori* from the physical information about them, then it seems like a sufficiently rich account of the physical information about human beings should be enough to deduce *a priori* all the truths about human beings. If nothing else, the burden of proof is placed on the anti-*a priori* physicalist to deny that physicalists are committed to either: (*a*) that all the truths about simple organisms can be deduced *a priori* from enough information about their physical nature and environment; (*b*) that human beings differ from simple organisms only by having more complex ingredients and arrangements; or (*c*) by virtue of having more complex ingredients and arrangements, there are new truths about human beings that cannot be *a priori*

---

[71] Jackson (1995), p. 415.

deduced from all the physical information about human beings. I find it difficult to see how a physicalist could provide a reason for rejecting (*a*), (*b*), or (*c*) without it being very implausible or *ad hoc*. At least given my understanding of physicalism, it follows that the PKI and psychophysical condition can be taken as *a priori* necessary claims.

1.2 Dualism

In this section I will spell out how to understand the sort of dualism that I intend to defend using the Knowledge Argument.[72] There are a variety of different kinds of dualisms and ways to characterize dual entities. The kind of dualism that the Knowledge Argument supports is property dualism, which should be understood differently from predicate dualism and substance dualism.

Predicate dualism is the view that physically irreducible psychological or mental predicates are necessary to give a complete description of the world.[73] The motivations for predicate dualism follow from recognizing that no Nagelian bridge laws can be given to account for reducing mental predicates to physical predicates as well as considering the multiple realizability of certain higher-order mental states. These reasons have been previously dealt with as motivations for non-reductive physicalism. Consequently, predicate dualism is a form of dualism in language alone. Predicate dualism makes no ontological

---

[72] I have phrased the sentence this way to leave open the possibility of other kinds of dualism being viable as a result of other arguments.

[73] See Robinson (2008), §2.1.

commitments, thereby leaving open the possibility of non-reductive physicalism. I have argued in previous sections that non-reductive physicalism is not a plausible way to be a physicalist, and for similar reasons it will follow that I do not take predicate dualism as a viable option for the thoroughgoing physicalist.

Ontological dualism, as opposed to linguistic dualism, recognizes that reality consists of two fundamentally different kinds of reality, typically where one sort of reality is fundamentally physical and the other is essentially not physical. Ontological dualism comes in at least two varieties, substance dualism and property dualism. Substance dualists maintain that the different kinds of things that constitute the world are two kinds of substances, physical and non-physical. Substance dualism is widely disparaged in some contemporary philosophical circles as intellectually and scientifically untenable. For example, Daniel Dennett has recently written:

> Dualism (the view that minds are composed of some nonphysical and utterly mysterious stuff) . . . [has] been relegated to the trash heap of history, along with alchemy and astrology. Unless you are also prepared to declare that the world is flat and the sun is a fiery chariot pulled by winged horses—unless, in other words, your defiance of modern science is quite complete—you won't find any place to stand and fight for these obsolete ideas.[74]

Despite this sort of rhetorical abuse leveled against substance dualism, the view still enjoys a large following among a wide array of philosophers and scientists

[74] Dennett (1996), p. 24.

today.[75] The Knowledge Argument, however, does not directly argue for the truth or falsity of substance dualism.[76] So, the details of how to understand substance dualism are immaterial to this current work.

Property dualism is another kind of ontological dualism. Unlike substance dualists who believe that there are physical and non-physical substances, property dualism is compatible with a worldview that takes all substances to be physical in nature. Property dualism differs from physicalism by maintaining that non-physical properties exist as fundamentally irreducible characteristics of the way the world is. Although any example of a non-physical property is going to be contentious, typically property dualists take the properties of conscious experience to be paradigmatic of the sort of properties that are introduced as novel and irreducible. Unlike predicate dualists who take the irreducibility of certain phenomena to be merely linguistic or epistemic, property dualists take the irreducibility of certain phenomena to be an ontological thesis. In other words, property dualists maintain that non-physical properties are needed to capture the structure of the world. The predicate dualist, on the other hand, is minimally committed to the thesis that our

---

[75] Such as Penfield (1975); Popper and Eccles (1977); Robinson (1982); Hart (1988); Foster (1991); Braine (1992); Taliaferro (1994); Stump (1995); Yandell (1995); Quinn (1997); Swinburne (1997); Hasker (1999); Zimmerman (2003); Dilley (2004); Goetz (2005); Plantinga (2006); Unger (2006); Moreland (2008); Beauregard and O'Leary (2008); Leftow (2010); Lowe (2010).

[76] Some take Swinburne's brain-splitting thought experiment to be a kind of Knowledge Argument for substance dualism. See Swinburne (1984), (1996), (1997), and (2009). See Melnyk (2003), pp. 178-180 (especially p. 179, n. 5) for a reading of Swinburne that classifies his argument as a Knowledge Argument akin to the famous arguments of Nagel and Jackson. I will not include Swinburne's argument for substance dualism among my consideration of the mainstream Knowledge Argument, although I am sympathetic to his argument.

language about the world will not admit a reductive analysis of certain mental properties.

It is difficult to give a precise characterization of non-physical properties. For example, it would be outright question-begging to define non-physical properties as mental properties since the burden is on the property dualist to show that mental properties are not physical properties. Some have suggested that non-physical properties should be characterized as possessing intentionality or "about-ness."[77] Stated roughly, the claim that non-physical mental states have intentionality can be thought of claiming that mental states are characterized by being about something. For example, Rick has *beliefs* about golf, or Diane *fears* that her cat might scratch her, or John *hopes* that his car is going to start. All of these mental states are characterized by intentionality; each mental state is about something. Once again, the stipulation that any system with intentionality is by definition a system with non-physical properties is question-begging since it remains to be shown that intentional features are not reducible to physical realization.[78]

I think it is best to leave the characterization of non-physical properties as a negative description. The property dualist is claiming that there are at least two fundamental kinds of properties. On the one hand, there are physical

---

[77] Such as Brentano (1874) and Chisholm (1967).

[78] Perhaps the most notable attempts to provide a physical account of intentionality are Dennett (1971) and Dretske (1995).

properties, which can be picked out by being either (i) part of the fundamental description of a completed physics, or (ii) functionally realized by systems that fundamentally consist of entities in a completed physics. On the other hand, there are non-physical properties, which are properties that ontologically are not part of the fundamental constituents of a completed physics and are not functionally realized by systems that fundamentally consist of entities in a completed physics. On this characterization, either physicalism or property dualism is true. If physicalism is false, then property dualism is true.

The Knowledge Argument sets out to show that physicalism is false and therefore property dualism is true. Since physicalism is committed to the *a priori* necessity of (PKI), the Knowledge Argument shows that there are some properties that we know exist whose existence cannot be inferred *a priori* from the complete set of physical truths. How we know that there are such properties and how we can conclude that they cannot be deduced *a priori* from the complete set of physical truths will be the task of the next chapter.

CHAPTER 2: THE KNOWLEDGE INTUITION, DIRECT ACQUAINTANCE, AND KNOWLEDGE OF QUALIA

Chapter 1 provided the conceptual distinctions that differentiate dualism and physicalism. On my account, physicalism amounts to the claim that all that exists is either a fundamental constituent of reality described by physics or causally realized by processes that are ultimately constituted by entities that are physical. The dualist rival is the position that not everything that exists is satisfied by physicalism's account of reality. As an important corollary to physicalism, I also identified the Physicalist Knowledge Intuition (PKI) as providing one condition whereby physicalism can be falsified. According to the PKI, if knowing all the physical truths is not sufficient for knowing all the truths about the world, then physicalism is false. What remains to be shown is whether there are any truths that we know which cannot be derived from the complete set of physical truths. In this chapter, I will present the *prima facie* case for the knowledge intuition (that there are truths which we know that cannot be derived from the complete set of physical truths), and then substantiate that insight by appealing to the notion of direct acquaintance. This chapter will complete my positive case for the Knowledge Argument (by providing positive arguments for accepting it), and the following three chapters will constitute a negative case for the Knowledge Argument (by responding to objections to the Knowledge Argument). Together, they present my overall defense of the Knowledge Argument.

2.1 Knowledge of Qualia

As I have previously argued, physicalism implies the PKI, which states a condition whereby physicalism is falsified. Physicalism is deemed to be false, according to the PKI, if all the physical information about the world is not sufficient to provide all the propositional knowledge about the world. The *prima facie* case that our knowledge of qualia—the subjective character of conscious experience—is the sort of knowledge that cannot be derived from physical information has been canonically stated by Frank Jackson's renowned thought experiment:

> 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 specializes 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 use terms like 'red', 'blue', and so on. . . .
>
> 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.[1]

The intuition that this thought experiment is supposed to motivate is clear enough. No matter how much physical information Mary possesses, she is never going to be in a position to figure out what it's like to be appeared to redly. In other words, physical information alone is not sufficiently robust to deduce the truths that characterize conscious experience. Stoljar and Nagasawa, whose

[1] Jackson (1982), p. 130. See also Jackson (1986), p. 291.

terminology I will follow, refer to this as the "knowledge intuition."[2] Perhaps visual experience is the most vivid illustration of the intuition, but the thought experiment could be modified to accommodate qualia that accompany auditory experience, gustatory experience, tactile experience, olfactory experience, or pain experience. Below, I will present my defense of the knowledge intuition after a brief perusal of some relevant examples from the early modern philosophers Leibniz, Locke, Berkeley, and Hume.

2.1.1 Early Modern Philosophers and the Knowledge Inutition

In his *Monadology*, Leibniz has a suggestive passage that points to the sort of insight that underwrites the knowledge intuition.

> [I]t must be confessed that perception and that which depends upon it are inexplicable on mechanical grounds, that is to say, by means of figures and motions. And supposing there were a machine, so constructed as to think, feel, and have perception, it might be conceived as increased in size, while keeping the same proportions, so that one might go into it as into a mill. That being so, we should, on examining its interior, find only parts which work one upon another, and never anything by which to explain a perception. Thus it is in a simple substance, and not in a compound or in a machine, that perception must be sought for.[3]

Although Leibniz employs his thought experiment to substantiate a point about the nature of simple substances, the basic idea is similar to the knowledge intuition. If Leibniz used "perception" to mean "the conscious experience of perception," then this passage can be read as an illustration of the knowledge

---

[2] Nagasawa and Stoljar (2004), pp. 2-5. Compare Alter (2006), §2.

[3] Leibniz (1714), para 17.

intuition. On this reading, Leibniz claimed that understanding the physical components of something with conscious experience is not sufficient to tell us about the intrinsic nature of its conscious experience.

Although the knowledge intuition does not depend on their strict form of empiricism, many of the British empiricists of the eighteenth century adhered to a similar intuition about one's ability to derive truths of mental experience from truths of the physical world. Consider, for example, John Locke's explanation of simple ideas:[4]

> Simple ideas . . . are only to be got by those impressions objects themselves make on our minds, by the proper inlets appointed to each sort. If they are not received this way, all the words in the world, made use of to explain or define any of their names, will never be able to produce in us the idea it stands for. For, words being sounds, can produce in us no other simple ideas than of those very sounds; nor excite any in us, but by that voluntary connexion which is known to be between them and those simple ideas which common use has made them the signs of. He that thinks otherwise, let him try if any words can give him the taste of a pine-apple, and make him have the true idea of the relish of that celebrated delicious fruit. So far as he is told it has a resemblance with any tastes whereof he has the ideas already in his memory, imprinted there by sensible objects, not strangers to his palate, so far may he approach that resemblance in his mind. But this is not giving us that idea by a definition, but exciting in us other simple ideas by their known names; which will be still very different from the true taste of that fruit itself.[5]

We can see from this passage that Locke believes that in order to acquire simple ideas, such as the taste of a pineapple or the look of a particular color or the

---

[4] Sometimes early modern philosophers refer to "ideas" as the most basic kinds of beliefs, and at other times they can mean for "ideas" to mean the sensations of conscious experience. In my discussion of the early moderns, I will use the term, "idea," for the belief state, and sensation for the conscious experience.

[5] Locke (1689), bk 3, ch 4, para 11.

auditory qualities of a particular sound, one must go through the process of having those sensations. Where Locke resonates with the knowledge intuition is in his affirmation that one cannot come to know these simple ideas through descriptions and definitions; one can only come to know them from one's firsthand sensations.[6] Locke continued in the passage quoted above to write that denying the knowledge intuition would be tantamount to accepting that a person who has been blind from birth could come to know the intrinsic character of color experiences from a rich enough description of the experience, which Locke takes to be manifestly absurd. As we shall see, this is a widely accepted intuition among the empiricists.

George Berkeley took much pleasure in arguing that the alleged physical descriptions of the causes of "sensible ideas" or sensations are insufficient to confer knowledge of the intrinsic nature of those ideas.[7] For example, in discussing our knowledge of sound, Berkeley rejects and ridicules the suggestion that sounds are nothing more than particles in motion that affect people's auditory nerves.[8] Berkeley is willing to mock anyone who contends that the real nature of sound is a certain kind of motion or vibration of molecules, since that

---

[6] See also, Locke (1689), bk 2, ch. 1, sec 6: "I think, it will be granted easily, that if a child were kept in a place where he never saw any other but black and white till he were a man, he would have no more ideas of scarlet or green, than he that from his childhood never tasted an oyster, or a pine-apple, has of those particular relishes."

[7] Of course, since Berkeley does not think that physical objects cause our sensations, his thesis is counterfactual. Perhaps, he would say something like this: Even if physical causes were the causes of our ideas about the intrinsic nature of sensations, knowing the physical causes of these ideas would not be sufficient for knowing the ideas they allegedly produce.

[8] Berkeley (1713), pp. 181-183.

person would be committed to the position that "real sounds may possibly be seen or felt, but never heard."[9] Berkeley provides similar arguments throughout the first part of his *Three Dialogues* to maintain that the immediate sensations of pain, pleasure, color, odor, and tactile feelings are different from their supposed causes. In conjunction with Berkeley's controversial thesis that "an idea can be like nothing but an idea,"[10] it follows that ideas about the intrinsic nature of sensations cannot be inferred from non-sensible ideas. Thus, Berkeley affirms the knowledge intuition. For this reason, Berkeley takes the example of a person's conception of colors who has been blind from birth as a paradigm of a person who is profoundly and helplessly ignorant about the intrinsic nature of colors.[11]

David Hume also agrees with the knowledge intuition. In defense of his empiricism he wrote,

> If it happen, from a defect of the organ that a man is not susceptible of any species of sensation, we always find that he is as little susceptible of the correspondent ideas. A blind man can form no notion of colours; a deaf man of sounds. Restore either of them that sense, in which he is deficient; by opening this new inlet for his sensations, you also open an inlet for the ideas; and he has no difficulty conceiving these objects.[12]

Hume's point accords with the knowledge intuition. No amount of descriptions or information can provide a congenitally blind person with the idea of color; nor

---

[9] Berkeley (1713), p. 182.

[10] Berkeley (1710), section 8.

[11] Berkeley (1710), section 77

[12] Hume (1748), sec 2, para 7.

can any descriptions give a congenitally deaf person the idea of sound. Like the other British empiricists, Hume claimed that in order to possess these ideas a person must experience the requisite sensation. The knowledge of these sensations cannot be acquired through other means.

2.1.2 Twentieth Century Examples of the Knowledge Intuition

A number of twentieth century philosophers have presented thought experiments that support the knowledge intuition. Consider, first, Bertrand Russell's passing comments on the difference between the motion of light particles and understanding the intrinsic nature of light:

> It is sometimes said that 'light *is* a form of wave-motion', but this is misleading, for the light which we immediately see, which we know directly by means of our senses, is *not* a form of wave-motion, but something quite different—something which we all know if we are not blind, though we cannot describe it so as to convey our knowledge to a man who is blind. A wave-motion, on the contrary, could quite well be described to a blind man, since he can acquire a knowledge of space by the sense of touch; and he can experience a wave-motion by a sea voyage almost as well as we can. But this, which a blind man can understand, is not what we mean by *light*: we mean by *light* just that which a blind man can never understand, and which we can never describe to him.[13]

Like the British empiricists before him, Russell contends that a congenitally blind man could not come to deduce what it's like to experience light visually from understanding the physical properties of light particles. While a blind man could come to understand perfectly the physics that underlie light particles, he would not come to know the intrinsic features of light that the

[13] Russell (1912), p. 28.

sighted understand immediately through visual experience. Russell affirms the knowledge intuition in his statement that our understanding of light is something "we can never describe to him [referring to a congenitally blind man]."

Another significant example of the knowledge intuition in the twentieth century is C. D. Broad's thought experiment concerning an "archangel"—a being that can infallibly perform any mathematical and logical deductions—who is given all the fundamental physical information about the world. Broad suggests that such a being would still not know everything that is true in the world:

> Take any ordinary statement, such as we find in chemistry books; e.g., "Nitrogen and Hydrogen combine when an electric discharge is passed through a mixture of the two. The resulting compound contains three atoms of Hydrogen to one of Nitrogen; it is a gas readily soluble in water, and possessed of a pungent and characteristic smell." If the mechanistic theory be true the archangel could deduce from his knowledge of the microscopic structure of atoms all these facts but the last. He would know exactly what the microscopic structure of ammonia must be; but he would be totally unable to predict that a substance with this structure must smell as ammonia does when it gets into the human nose. The utmost that he could predict on this subject would be that certain changes would take place in the mucous membrane, the olfactory nerves and so on. But he could not possibly know that these changes would be accompanied by the appearance of a smell in general or of the peculiar smell of ammonia in particular, unless someone told him so or he had smelled it for himself.[14]

Broad's archangel motivates the knowledge intuition by trying to show that a being like the archangel would know all the physical information about the world, but it would lack knowledge of the intrinsic qualities of sensations, like the specific character of odor that ammonia possesses. Unless the archangel

[14] Broad (1925), p. 71.

had experienced the smell for itself or had been told by someone that ammonia had a smell, Broad claimed that a being like the archangel would not even know that there is such a thing as the sensation of smell. Broad's archangel is quite similar to Jackson's Mary: the archangel possesses all the physical information about the world and yet the archangel fails to know something about our conscious experiences of the world.

In his renowned article, "What is it Like to be a Bat?" Thomas Nagel provides another example that has been used to pump the knowledge intuition.[15] His thought experiment invites his readers to imagine that they have a complete understanding of the physiology of bats. From that information, however, it seems clear that it would not be sufficient to figure out what it's like to undergo the conscious experiences of a bat. Taking his cue from Nagel, Laurence BonJour has developed Nagel's thought experiment to consider whether bat-like aliens could come to know what it's like to undergo the conscious experience of a human being from understanding completely the physiology of human beings.[16] Since the bat-like aliens would not have conscious experiences remotely analogous to the modalities whereby humans undergo conscious experience, it seems like there is no way that the bat-like aliens could come to infer what it's like to be human (for the same reasons that humans cannot know what it's like to be a bat). If the knowledge intuition were false, then we would expect as a

[15] Nagel (1974).

[16] BonJour (2005).

matter of course that humans could infer what it's like to be a bat and that bat-like aliens could infer what it's like to be a human from a sufficiently rich pool of physical information. But it is hard to see how either of these inferences could be achieved. Thus, these thought experiments support the knowledge intuition.

2.1.3 A *Prima Facie* Justification of the Knowledge Intuition

From the brief sketch presented above, we can see that there is some historical precedent for the knowledge intuition. If both the PKI and the knowledge intuition are correct, then we can infer that physicalism is false. The argument would follow from these premises:

(PKI) If complete possession of all knowledge of physical truths isn't sufficient to provide all propositional knowledge of the actual world, then physicalism is false.

(KI) Complete possession of all knowledge of physical truths isn't sufficient to provide all propositional knowledge of the actual world, namely knowledge of the subjective character of conscious experience.

Therefore,

(¬P) Physicalism is false.

Since I have defended my reasons for accepting (PKI) in chapter 1, all that remains for my positive case of the Knowledge Argument is justifying my reasons for accepting (KI). Is there more that can be said about (KI) besides the intuitive thought experiments that have been cataloged above? As we will see in chapters 4 and 5, it is possible for physicalists to acknowledge that Mary learns

something new, but to deny that she comes to have new propositional knowledge or knowledge of a new truth. Prior to examining these alternative accounts of Mary's new knowledge more closely, it would be helpful to motivate the position that Mary's new knowledge is propositional or factual knowledge. What more can be said to substantiate that the knowledge intuition supports the claim that having all the physical information about the world is not sufficient to possess all the propositional knowledge about the world?

Before defending an account of knowledge based on direct acquaintance to substantiate the knowledge intuition, we can consider another thought experiment that underwrites the factual or propositional nature of the knowledge intuition. Martine Nida-Rümelin presents a thought experiment using Marianna, a woman who has never experienced color (for whatever reasons), although her vision is normal and she believes so justifiably.[17] Marianna agrees to participate in a psychological experiment where she is placed in a house where everything is colorfully and arbitrarily decorated, which is where she experiences color for the first time. Marianna, however, is not taught the names of these colors, nor is she allowed to see objects that she knows to be a certain color (from testimony alone), such as that the sky is blue, ripe tomatoes are red, bananas are yellow, grass is green, etc. At the end of the experiment, she is presented with four slides of clear examples of red, yellow, blue, and green, and she is asked which color sample she thinks corresponds with the color

[17] Nida-Rümelin (1995).

normal sighted people see when they look at the sky (under ordinary daytime conditions). We can imagine that after deliberating about the descriptions she has heard of the beauty of the clear blue sky, Marianna points to the red slide and proudly says, "I believe it is this one." Of course, she is wrong. It is not until Marianna leaves the psychological experiment that she finally comes to know the truth that the sky appears to be blue in the phenomenal sense.[18]

According to Nida-Rümelin, the point of the Marianna example is to show that

> Mary and Marianna acquire a particular kind of belief *that* the sky appears blue to normal perceivers, namely the phenomenal belief that it appears blue to normal perceivers, where phenomenal belief involves the application of the appropriate phenomenal concept. Both may have believed, in a sense (the non-phenomenal sense that does not require use of phenomenal concepts) that the sky appears blue to normal perceivers while still in their black-and-white environment (they may have been told so by their friends).[19]

From this insight we can conclude that both Mary and Marianna don't merely come to know *this is what it's like to experience redness* for the first time. What is more important is that both Mary and Marianna come to know new propositional truths about phenomenal properties. Minimally they come to know the truth that phenomenal redness is exemplified. When Mary and Marianna form beliefs about the color of the sky before they have had any color experience, their beliefs are composed of non-phenomenal color concepts—the

---

[18] It is possible to combine the Mary and Marianna thought experiments. For a recent example of this, see BonJour (2010), pp. 10-11.

[19] Nida-Rümelin (2008), §3.3.

kind of concepts about color that congenitally blind people have (e.g., "blue" means the color, whatever it is, that people refer to when they describe the color of the sky, etc.). When Marianna believes that the red slide corresponds to the color of the sky, it shows that even though she has the concept of phenomenal blueness, she lacks knowledge as to which phenomenal concept corresponds to the way the sky appears.[20] Finally, when both Mary and Marianna see the blue sky for the first time, they come to know for the first time that the sky appears to be phenomenally blue.

In addition to knowing that phenomenal properties are exemplified and the truths about one's perception of the world that involve phenomenal concepts, there are a number of other truths that Mary comes to know when she is released from her black-and-white environment.[21] For example, Mary knows for the first time the following modal truths about redness itself: that necessarily red is a color; that necessarily something cannot be red and green all over at the same time; that necessarily red is darker than yellow. Mary also comes to know some new truths about her sensations. For instance, Mary comes to know that sensations of red are more like sensations of green than they are like sensations

[20] If we also imagine that Marianna knows the complete set of physical truths and their deductive implications, it seems to do no help in aiding her in matching the phenomenal property with the object it is associated. On this point, see BonJour (2010).

[21] Mary's knowledge of these facts was suggested to me by Moreland (2003) and (2008).

of sourness. Intuitively, Mary has no grasp of these truths until she has the relevant experiences.[22]

I think it is fair to say that both Mary and Marianna present a *prima facie* case for accepting the key parts of the knowledge intuition. Even John Searle, who is no friend of dualism, agrees with the knowledge intuition when he describes Jackson's argument as "ludicrously simple and quite decisive."[23] First, the Mary thought experiment pumps the intuition that knowing all the physical information about the world does not inform us of all the knowledge there is about the world. Second, the Marianna thought experiment supports the intuition that the ignorance shared by both Mary and Marianna is ignorance of a truth. Thus, these thought experiments constitute a *prima facie* case for knowledge intuition. A more substantial defense of the knowledge intuition can be mounted, but it requires an explanation and defense of a controversial account of empirical knowledge. In the next section, I provide my account of empirical knowledge based on direct acquaintance. In the final section of this chapter, I will argue that direct acquaintance secures the insight of the knowledge intuition.

---

[22] One might argue that Mary could come to know these truths by testimony. However, it is controversial to claim that necessary truths can be known as necessary truths via testimony. Moreover, even if Mary can come to know by testimony the necessary truths in this paragraph, it is important to realize that Mary still wouldn't understand the propositions that express these necessary truths. After she experiences phenomenal redness, Mary is in a position to "see" the modal status of these propositions because of their content.

[23] Searle (1992), p. 118.

2.2 Direct Acquaintance

My substantial defense of the knowledge intuition builds on an epistemological view that relies on direct acquaintance for the foundations of empirical knowledge. In this section I will present and defend my account of direct acquaintance. In the following section I will show how an epistemology built upon direct acquaintance supports the knowledge intuition.

The concept of acquaintance was introduced to contemporary philosophy by Bertrand Russell in his article "Knowledge by Acquaintance and Knowledge by Description" and in chapter five of *The Problems of Philosophy*.[24] Russell explains that a person is acquainted with an object when he stands in a "direct cognitive relation to the object."[25] In another place, he writes "we have *acquaintance* with anything of which we are directly aware, without the intermediary of any process of inference or any knowledge of truths."[26] Russell's account of direct acquaintance has been co-opted by many advocates of classical foundationalism in the past century, which has resulted in some more nuanced accounts of the notion of acquaintance (hereafter, I will use the terms "direct acquaintance" and "acquaintance" interchangeably).[27] My understanding of

---

[24] Russell (1910); (1912).

[25] Russell (1910), p. 108.

[26] Russell (1912), p. 46.

[27] Some of the reasons for this maneuver have sketched in Fumerton (2008). Some notable examples of recent classical foundationalists who make use of acquaintance include Russell (1912), Lewis (1929), Price (1950), Fumerton (1995), Fales (1996), and BonJour (2003).

acquaintance will closely follow the account given by Richard Fumerton,[28] although there will also be some significant differences.

[28] Especially, in Fumerton (1995), pp. 73-85, (1998), and (2008).

2.2.1 An Account of Acquaintance

As I understand it, direct acquaintance is a simple, unanalyzable relation that holds between a mind and various entities, such as certain kinds of properties, relations, or facts. Since acquaintance is *sui generis* and resists being analyzed into simpler parts or classified under other kinds of relations, acquaintance must be taken as a primitive relation or concept. Acquaintance is not essentially an *intentional state* since it is possible to be directly acquainted with something without representing the object of that acquaintance with an intentional mental state. For example, for one to be acquainted with the fact that *p*, it does not necessarily imply that one has the belief that *p*; in other words, it is possible to be acquainted with the fact that *p* and simultaneously not have the belief that *p* (or any other intentional state that *p*). Furthermore, direct acquaintance is not in and of itself essentially an epistemic relation. While multiple instances of acquaintance can constitute a person's having justification or knowledge, this should not lead us to conclude that a single instance of acquaintance necessarily carries any epistemic force.

As tends to be the case with primitive concepts, the best one can do, perhaps, is to gesture at certain analogies, metaphors, and examples to help point others to the basic concept one is trying to pick out. One must keep in mind, however, that any of these examples taken too strictly could be misleading. For example, one might suggest a spatial metaphor to capture direct acquaintance—a person is directly acquainted with something when there is nothing *in between*

the person and the thing with which one is acquainted. Sometimes acquaintance is described as the relation that holds when something is immediately *before* one's mind. Typical examples of acquaintance include one's immediate awareness of being in pain such as one's awareness of one's own throbbing pain. One's awareness of this kind of pain is unmediated, not inferred through intermediary thoughts or experiences, and it is directly present to one's mind. Most classical foundationalists who make use of acquaintance believe that one can be acquainted with one's sensory states of mind, some of one's own beliefs, certain relations that hold between states of mind such as correspondence and entailing relations, and more controversially some acquaintance theorists have held that one can be acquainted with universals and the concept of causation.[29] Although I find it very plausible to think we can be acquainted with many more things, it is sufficient for this project to maintain that we can be acquainted only with some of our own sensory states of mind (as complexes of properties or facts), some of our own beliefs, and the relation of correspondence that can hold between objects of one's acquaintance.

Finally, most acquaintance theorists have held that one can be acquainted with some of one's own instances of acquaintance. While this is the most decisive evidence for believing that acquaintance relations exist, it can be useless as an attempt to convince those who are skeptical of acknowledging

---

[29] Russell (1912), pp. 51-52 claimed that we are directly acquainted with universals. Fales (1990) contends that we are directly acquainted with universals and causation.

acquaintance. For those who cannot "see" that they are capable of being directly acquainted with anything, maybe the best one can do as a final attempt to show the plausibility of acquaintance is to illustrate that denying acquaintance results in an epistemic disaster. One way to do this, put crudely, is to argue that if one does not ultimately ground one's justification in some direct acquaintance with a truth, then the possibility of vicious regresses and other epistemic disasters are lurking in one's account of justification. This is sometimes called the regress argument, and if successful, it establishes a type of foundationalism.[30] Foundationalism is the position that in order for some beliefs to be justified, there must be some beliefs that are epistemically basic or non-inferentially justified. Why we need non-inferentially justified beliefs and how direct acquaintance provides a non-arbitrary way to acquire them will constitute my last line of argument for direct acquaintance.

As it turns out, there isn't just one regress argument for the kind of foundationalism that is based on direct acquaintance—there are two.[31] First there is an epistemic regress. Suppose that contrary to foundationalism, in order for someone to be justified in believing any proposition *P*, one must *always* be inferentially justified in believing *P*, which is to say that one must be justified in believing that some set of reasons, R_1, makes *P* very probable. But since all

---

[30] For some recent versions of the regress argument, see Moser (1989), ch. 4; Fumerton (1995), pp. 55-60, 89-92; McGrew (1995), ch. 3; Audi (2003), ch. 7; Fumerton (2006a), pp. 40-42.

[31] Following Fumerton (1995) and (2006a).

justification would need to be inferential if foundationalism is false, then in order for a person to be justified in believing that R_1 make P very probable, it would require the person to have some other set of reasons R_2 from which the person legitimately infers that R_1 is likely to be true. But now, those who reject foundationalism will need to infer R_2 from a set of reasons R_3 in order for R_2 to be inferentially justified. It should be apparent that this regress is going to continue without end. Thus, if foundationalism is false and all justification is inferential, then, in order to be justified in believing anything at all, it would require having an infinite number of justified beliefs (which no mortal can accomplish). But since we are clearly justified in having some beliefs and we are not capable of having an infinite number of justified beliefs, it follows that those who reject foundationalism are mistaken. Hence, some beliefs are justified non-inferentially and foundationalism is correct.

The second regress for foundationalism is a conceptual regress. Suppose a non-foundationalist in providing a conceptual analysis of what justification means suggests that what it means for a person to be justified in believing some proposition is for the person to infer it from some other proposition (or set of propositions) that the person justifiably believes. The problem with this analysis of the meaning of justification is that it uses the concept of justification in the process of providing the meaning of justification. Thus, to understand the non-foundationalist concept of justification, a person must presuppose an understanding of the concept of justification. So, a non-foundationalist analysis

of the concept of justification results in a circular analysis. Fumerton compares this conceptual regress to an obviously flawed analysis of goodness where one tries to analyze all goodness in terms of instrumental goodness.[32] All goodness cannot be instrumental, claims Fumerton, because we could not find the origin of the source of goodness—something which has intrinsic goodness. Likewise, the concept of justification cannot be given an informative analysis if the meaning of justification requires all justification to be inferential. If all justification were inferential, we would never come to understand the meaning of justification. Just as instrumental goodness requires an understanding of intrinsic goodness, so understanding inferentially justified beliefs requires an understanding of non-inferentially justified beliefs.

The regress arguments demonstrate that giving up foundationalism has dire consequences. But how do they support the kind of foundationalism that I am offering, which is based on direct acquaintance? There are at least three reasons why the regress arguments support a foundationalism based on direct acquaintance. First, direct acquaintance provides a plausible way for a person to have a non-inferentially justified belief without giving up the notion that justification involves the ability to "see" from one's perspective that a belief is justified.[33] Second, on the account that I will provide below, it stops the

---

[32] Fumerton (1995), pp. 89-90, Fumerton (2006a), p. 41.

[33] I will take it for granted that the concept of justification is an internalist concept. Internal justification, on my view, requires as a necessary condition for being justified that from the subject's perspective, the subject can "see" the positive epistemic status of the justified belief.

epistemic regress by having the subject directly acquainted with everything one needs to know a *truth*. Third, since direct acquaintance is not an epistemic relation, analyzing justification or knowledge in terms of direct acquaintance will avoid the problem of creating a regress with one's concept of justification. Below, I will provide my account of non-inferential justification, where I will make good on these claims.

2.2.2 Acquaintance and Non-Inferential Knowledge

Although direct acquaintance is not essentially an epistemic state, this does not preclude the possibility that multiple instantiations of acquaintance can constitute an epistemic state such as having knowledge or justification. In fact, if epistemic concepts are not ultimately analyzed in terms of non-epistemic concepts, then the analysis of epistemic concepts is likely to be circular, as we saw in the conceptual regress argument above. Following Fumerton,[34] I will maintain that non-inferential empirical knowledge consists of the subject's having three kinds of acquaintance. The analysis of non-inferential knowledge I endorse claims that a person *S* has non-inferential empirical knowledge that *p* if and only if (i) *S* is directly acquainted with the belief that *p*; (ii) *S* is directly acquainted with the fact that *p*; and (iii) *S* is directly acquainted with the

[34] Fumerton (1995), especially pp. 73-85.

correspondence that holds between the fact that *p* and the belief that *p*.[35]

It is important to notice that this analysis of non-inferential knowledge does not require for the knowing subject to have and employ the concept of direct acquaintance in order to have knowledge. For example, a person can satisfy condition (i) without having the concepts of acquaintance or belief. Likewise, conditions (ii) and (iii) do not imply that the knowing subject understands the concepts of acquaintance or correspondence. Since acquaintance is not an intentional or epistemic state, a person can be acquainted with *p* without believing that she is acquainted with *p*. As we've seen, placing a requirement of this sort on knowledge (or justification) results in a vicious infinite regress, and therefore it is a virtue of this account of knowledge that it blocks this epistemic regress.

The account of non-inferential empirical knowledge that is on offer here claims that in order for *S* to know that *p*, *S* must be acquainted with the belief that *p*, the fact that *p*, and the correspondence that holds between the belief and the fact that *p*. These three instances of acquaintance secure everything that the subject needs in order to have a truth.[36] On typical correspondence theories of truth, a truth-bearer (such as a thought or proposition) is true when and only when it corresponds to a truth-maker (such as a fact or states of affairs) and false

---

[35] I have sidestepped the difficult issue as to whether one must also base one's belief on these three acquaintances to count as having knowledge. I will continue on the assumption that if there is a well-defined understanding of the basing relation that it can be satisfied on my account without any problems.

[36] At least for the subject to secure a truth on a correspondence theory of truth.

otherwise. Therefore, when a subject has the three states of acquaintance needed to have non-inferential empirical knowledge on the present account, then he is directly acquainted with everything that he needs to know a truth: a truth-bearer (the belief that *p*), a truth-maker (the fact that *p*), and the correspondence that holds between the truth-bearer and the truth-maker (the correspondence between the belief that *p* and the fact *p*). When a person has these three acquaintances, Fumerton observes, "there is nothing more that one could want or need to justify a belief."[37]

In order to understand better the account of non-inferential knowledge that I am proposing, and also to rebut some well-known objections, I will consider some criticisms of my position next. After dealing with these clarifications and objections, I will use this account of non-inferential knowledge to support the key premise in the Knowledge Argument—the knowledge intuition.

2.2.3 Objections to the Acquaintance Account of Knowledge

To help show the plausibility of using direct acquaintance to account for non-inferential empirical knowledge, I will consider and respond to some of the most significant problems that have been raised against accounts of this sort. First, I will consider a dilemma about propositional content that is often attributed to Wilfrid Sellars, although I will primarily follow how the dilemma

[37] Fumerton (1995), p. 75.

has been stated by Laurence BonJour. The second objection comes from Timothy Williamson's anti-luminosity argument, which some critics think can be used to undermine the first-person authority that seems to secure knowledge on the acquaintance account. Finally, I will address recent attempts to revive the speckled hen problem for direct acquaintance—a problem which aims to show that acquaintance with a truth-maker is not sufficient to confer non-inferential knowledge.

There is a well-known dilemma for foundationalists that is often attributed to Wilfrid Sellars,[38] but which has been most clearly articulated by Laurence BonJour prior to his conversion to classical foundationalism.[39] To understand the dilemma it is necessary to elucidate the notions of conceptual and propositional content. Something has conceptual content when in order for a subject to be aware of it (or form a belief about it), it must fall under categories of thought. Something has a propositional nature when it can take a truth-value. Beliefs[40] are typically thought to have a propositional and conceptual character since beliefs structurally are composed of concepts, and beliefs can take a truth-value. Perhaps any putative example of a mental state that is non-conceptual and non-propositional will be controversial, but one plausible instance of

---

[38] Sellars (1956).

[39] BonJour (1985), p. 27. For another recent statement of this problem, see Williams (1999).

[40] Here I am not presupposing a view about the fundamental bearers of truth. In other words, one could hold that beliefs have propositional content either because they are the primary bearers of truth or because they derive their truth-value from more fundamental truth-bearers, such as propositions.

something that is non-conceptual and non-propositional will be one's "raw" sensation of searing pain. When one is aware of searing pain (as the pain *itself*, not as the belief about the searing pain), the person's raw sensation is neither conceptual nor is it propositional. The searing pain just *is*, it isn't conceived as being a certain way. Unlike one's belief that he is in the state of searing pain, the searing pain itself cannot take a truth-value. For reasons that generalize from the example of searing pain, it could be argued that all of one's relevantly similar sensations are non-conceptual and non-propositional.

The dilemma raised by Sellars and BonJour exploits the divide between the propositional and conceptual nature of one's most basic empirical beliefs and the non-propositional and non-conceptual nature of the conscious experiential states that foundationalists claim underwrite those basic beliefs. If, on the one hand, the beliefs are propositional and conceptual and the experiences are non-conceptual and non-propositional, then it seems mysterious (if not outright impossible) to explain how beliefs can derive any justification from non-propositional and non-conceptual entities. The intuitive idea behind this horn of the dilemma has been expressed by Donald Davidson, who wrote, "nothing can count as a reason for holding a belief except another belief."[41] The basic point is that only conceptual and propositional entities can stand in logical relations to other conceptual and propositional entities. Since mental experiences lack conceptual and propositional features, they lack the ability to justify beliefs with

[41] Davidson (1986), p. 126.

propositional and conceptual content.

On the other hand, if one accepts that mental experiences are conceptual and propositional, then the problem for foundationalists is to explain how one is justified in accepting the conceptual and propositional content of experiences. An anti-foundationalist may stress that this horn of the dilemma pushes the problem of generating non-inferential justification back a step. If beliefs need justification in order for one to be justified in accepting their conceptual and propositional content, then experiences too need some justification for one to be justified in accepting their conceptual and propositional content. Contrary to foundationalism, this horn of the dilemma essentially states that mental experiences do not constitute a non-arbitrary way to stop the regress of justification needed to arrive at a foundational bedrock for empirical knowledge.

The acquaintance account of non-inferential knowledge that I am endorsing succeeds in averting the BonJour/Sellars dilemma essentially by grasping the horn of the dilemma that allows non-inferentially justified beliefs to derive their justification from sources that are non-conceptual and non-propositional in nature. The key to this solution is evident by recognizing that propositional and conceptual entities (such as beliefs) can *correspond* to non-propositional and non-conceptual entities (such as mental experiences). Correspondence, like acquaintance, is a *sui generis* relation, and thereby resists being analyzed in terms of other philosophical concepts. Thus, like acquaintance, correspondence is a basic relation that is difficult for one to argue

for because it is so basic. Some examples may help illustrate what the correspondence relation is. For example, one's non-propositional, non-conceptual mental experience of a red circle corresponds with the proposition that *I am being appeared to in a red circular way*, and not *I am being appeared to in a green triangular way*. Similarly, one's non-propositional, non-conceptual mental experience of seeming to see a cat on a mat corresponds to the proposition that *it appears to me that the cat is on the mat*, but it does not correspond to the proposition that *it appears to me that the mat is on the cat,* nor does it correspond to the claim that *it appears to me that the planet Venus is sour*. Of course, those who are skeptical about the relation of correspondence will undoubtedly find these examples question begging.

Perhaps objectors to this response are skeptical because they cannot understand precisely why certain non-conceptual, non-propositional entities can stand in correspondence relations to conceptual, propositional entities. While they grasp the examples noted above, they will likely conclude that conscious experiences must be propositional and conceptual since certain experiences clearly do correspond with certain propositions and not others. After all, objectors might claim, if mental experiences were totally devoid of all conceptual and propositional content, then there would be no reason to think that certain propositions correspond to the mental experience and others don't.

In response to this concern, I think some progress can be made. While one's given mental experience is not propositional or conceptual, this is not

tantamount to claiming that what is given in one's raw mental experience is completely unstructured and unintelligible. Following a suggestive passage from Evan Fales,[42] it may be helpful to think of mental experiences as being *protopropositional*. The exemplification of properties that constitute mental experiences will have an intrinsic structure that allows the right propositional and conceptual content to "map onto" the mental experience. In this way, mental experience can simultaneously be non-propositional and non-conceptual, while being sufficiently rich and discriminating for the right propositional and conceptual content to correspond to it.[43] By virtue of the protopropositional features of mental experiences (e.g., the intrinsic structure of the properties exemplified in a particular mental experience), we can understand how some propositions correspond (or fail to correspond) to certain mental experiences.

So, the acquaintance theorist can meet the Sellars/BonJour dilemma by accepting that non-conceptual and non-propositional entities can stand in correspondence relations to entities that are conceptual and propositional. Furthermore, by recognizing the protopropositional structure of mental experience, I have suggested a way in which mental experiences can correspond (or not) to basic empirical beliefs. Thus, the Sellars/BonJour dilemma can be averted by accepting one horn of the dilemma and rebutting the alleged problem

---

[42] Fales (1996), p. 169.

[43] As Fales (1996), p. 169 notes, the protopropositional features of mental experiences will typically be complex and outstrip the propositional content of any belief that corresponds to it.

raised by anti-acquaintance critics.

A second problem that I will consider challenges the notion of acquaintance by calling into question whether acquaintance can serve as a luminous "cognitive home" for knowledge. Luminosity is a property of a mental state such that if a mental state is luminous, then whenever a person is in that state she is in a position to know infallibly that she is in that state.[44] Timothy Williamson has argued in *Knowledge and its Limits* that there is no privileged state of mind that is able to confer knowledge infallibly, including acquaintance.[45] Since acquaintance is susceptible to fallibility that one may not be able to detect or correct, it cannot serve as an authoritative way to justify beliefs.

Williamson's argument begins by taking a paradigm state of mind that some philosophers have taken to be luminous, such as feeling cold.[46] If the state of feeling cold is luminous, then the person who is feeling cold will be in a position to know that he is feeling cold. However, Williamson imagines what happens throughout the day as the person slowly begins to warm up. Although he may not detect any change in his phenomenal experience from one moment to the next, over a long stretch of time he will cease to feel cold. The problem, according to Williamson, is that at some midpoint in the warming-up process he will feel cold but not be in a position to know that he is cold. Since he is unable

---

[44] Williamson (2000), p. 95.

[45] Williamson (2000), ch. 4.

[46] Williamson (2000), pp. 96-98.

to discriminate one minuscule change from one moment to the next, there will come a point where the subject feels cold at one moment but not feel cold at the next—yet he will be unable to distinguish one moment from the next. In the borderline cases where feeling cold is very close to not feeling cold the belief is "unsafe" because one could be in a very similar state (one that is virtually indistinguishable from it) and form a false belief. Due to the failure of safety, then, one could feel cold in the borderline cases and fail to know that one is feeling cold. Therefore, Williamson concludes, feeling cold is not luminous.

Next, Williamson generalizes his argument—what is true of cold can be true of any significant mental state.[47] His point is that any significant mental state occurs on a spectrum with imperceptible degrees, so there is no privileged mental state such that whenever a person is in that state, then, the person is in a position to know he is in it.

With regard to direct acquaintance, Williamson's anti-luminosity argument can be seen as a challenge to justify acquaintance as a reliable or distinctive source for non-inferential justification or knowledge. Given its vulnerability to error, why should one rely on acquaintance for one's foundational justification, rather than another source? Some may even be concerned that if acquaintance cannot guarantee one's knowledge of being in a mental state such as feeling cold, then it cannot be trusted to ground any basic knowledge.

---

[47] Williamson (2000), pp. 106-109.

The first step for the acquaintance theorist to overcome Williamson's anti-luminosity argument is to point out that Williamson's concept of a luminous state of mind does not capture what acquaintance theorists have in mind when they have proposed that there are privileged states of mind that are more basic and epistemically secure than others. As Williamson defines a luminous state of mind, it is one such that whenever one is in that state, one is in a position to know that one is in it. But the motivation for relying on acquaintance for one's cognitive home is not that acquaintance is an infallible guide for knowing when a subject is in the state of acquaintance. As I emphasized above, non-inferential justification by acquaintance does not require for the subject to be justified in believing that one is in any state of acquaintance. Rather, the point is that acquaintance is the starting point of empirical knowledge because it is the best source of non-inferential justification, even if it isn't always an infallible source for determining whether the subject is in the state of direct acquaintance. In a recent article, Fumerton explains it this way:

> Our inner mental life constitutes a cognitive home not because we always have unproblematic access to our mental states. Our inner mental life constitutes a cognitive home because we sometimes have a kind of justification for believing truths about such states that is better than the justification we have for believing other empirical truths, and that is, in fact, as good as justification gets.[48]

Other critics of Williamson, have noted that Williamson's anti-luminosity argument only applies to boundary cases where knowledge fails because one is

[48] Fumerton (2009), p. 73.

not in a paradigmatic mental state.[49] Given the right specifications that rule out cases that involve non-paradigmatic mental states, one can specify a subset of mental states that are luminous, thereby averting Williamson's argument. Perhaps then, the force of Williamson's argument can be taken to challenge that one can never *really* be sure that one has knowledge only when one is in a borderline mental state. First, a person's inability to discriminate when she is in a non-paradigmatic state is no reason to doubt that she has knowledge when she is acquainted with exemplary mental states. And second, even when borderline states of mind may rob the subject of her assurance that she has knowledge, this does not necessarily imply that she does not have knowledge.[50] Given my analysis of knowledge, it does not follow that whenever a person knows that *p*, then, she is in a position to know that she knows that *p*.

A more controversial, independent reason to reject Williamson's anti-luminosity argument is because there are mental states that infallibly put the subject in a position to have knowledge.[51] Consider the case where a person forms a belief by referring directly to a mental state with which she is directly acquainted; typically we express these beliefs as "I am experiencing *thusly*." This process of belief formation is guaranteed to be infallible because, as Timothy McGrew explains, "it is literally impossible to form the belief that one is

---

[49] Such as Conee (2005); Hawthorne (2005); and Reed (2006).

[50] Fumerton (2001), p.15: "From the fact that a certain justification is infallible, it does *not* follow that one could not mistakenly believe that one has an infallibly justified belief."

[51] This response to Williamson is inspired by McGrew and McGrew (2007), pp. 118-127.

appeared to like *that* without there being a 'that' to refer to."[52] Since the mental state which is being referred to is partly constitutive of the very belief,[53] referentially formed beliefs cannot fail to be true because of the way such beliefs are formed. In other words, there is no possible way to form such a belief without its being true.

In response to this last concern, Williamson is willing to concede that referentially formed beliefs are luminous, but he thinks that such beliefs are trivially luminous and thereby they "constitute a very minor limitation on the generality of the argument."[54] To the contrary, it is not sufficient to brush aside this infallible source of empirical justification as trivial. The reason why referentially formed beliefs are not a trivial source of infallible justification is that referentially formed beliefs of this sort serve as the basis for a certain view of the meaning of empirical concepts. Given a certain type of internalism about meaning, direct acquaintance and reference supply epistemic subjects with the meaning of certain empirical concepts.[55] As will become evident in §5.2, I accept an account of meaning that is congenial to this approach. It is through one's direct apprehension of properties that the subject grasps the meaning of certain empirical concepts, like feeling cold or painfulness, and for this reason the

---

[52] McGrew (1995), p. 115. For similar accounts of infallible justifying processes see Moser (1985), pp. 173-187, Fumerton (1995), pp. 71-73, 76-77, and Fales (1996), pp. 143-155.

[53] Here is one place where my account of direct acquaintance significantly departs from Fumerton's.

[54] Williamson (2000), p. 109.

[55] For one recent example of this approach, see McGrew and McGrew (2007), pp. 122-125.

subject who is acquainted or referring to the property directly cannot be wrong.[56] Where fallibility can occur is not at the most basic level of belief, but in the ascent from their foundations, such as finding a public description to convey the character of one's belief formed by direct acquaintance or reference; or inferring the causes or conditions that brought about the basic belief; or in comparing the basic belief to other beliefs on the basis of memory. But these criticisms essentially concede that the foundations grounded in direct acquaintance and reference are secure.[57] Therefore, Williamson's anti-luminosity argument does not present an insuperable problem for a foundationalism based on direct acquaintance.

The third objection to using direct acquaintance as a basis for non-inferential justification is known as the problem of the speckled hen. An influential version of the problem was first stated by Gilbert Ryle,[58] and it has returned recently as a criticism of direct acquaintance from critics, including Peter Markie and Ernest Sosa.[59] The objection intends to show that being directly acquainted with a truth-maker is not sufficient to put one in a position to know a truth. For example, suppose a person has the appearance of a hen with 55

---

[56] I am presuming that semantic externalism is false. See §5.2 for my reasons to reject semantic externalism.

[57] A complete response to these objections would be necessary in order to answer epistemic puzzles such as skepticism about the external world. I believe these objections can be met, but to pursue them would be a significant detour in my current project.

[58] Ryle (1949), ch. 6.

[59] Markie (2009); Sosa (2003a), (2003b).

speckles and that the person is directly acquainted with his appearance. Yet, it seems that the person is not non-inferentially justified in believing that he is having an appearance of hen with exactly 55 speckles, even if he luckily forms the belief that he is having the appearance of a hen with 55 speckles. In contrast, a person who is directly acquainted with a simple image—such as three large red circles on a white background—surely is non-inferentially justified in believing that he appears to see three red circles on the basis of his direct acquaintance with the experience. Thus, the challenge, as Sosa puts it, is for the acquaintance theorist to "tell us *which* sorts of features of our states of consciousness are epistemically effective ones, the ones such that *by corresponding to them specifically* that our basic beliefs acquire epistemically foundational status."[60] The concern, of course, is that the acquaintance theorist has no principled way of explaining why one's acquaintance with a simple mental state is able to become a non-inferentially justified belief, whereas one's acquaintance with a complex mental state fails to result in a non-inferentially justified belief.

In response to the problem of the speckled hen, advocates of direct acquaintance have a number of ways to respond.[61] In what follows, I will provide the two best responses to the problem of the speckled hen, which should make it sufficiently clear that the speckled hen does not pose insurmountable problems for employing direct acquaintance to justify beliefs non-inferentially.

---

[60] Sosa (2003a), pp. 277-278. Similar remarks appear in Sosa (2003b), p. 121.

[61] See Fumerton (2005) for a helpful overview.

At the outset, it is important to keep in mind that the account of non-inferential justification being defended in this project does not take direct acquaintance with a truth-maker for a belief to be *sufficient* to have a non-inferentially justified belief. So, the problem of the speckled hen should not be construed in such a way that it presupposes that direct acquaintance with a truth-maker is sufficient for having a non-inferentially justified belief for whatever corresponds to the truth-maker. Recall that non-inferential justification requires three acquaintances: acquaintance with a belief, acquaintance with the truth-maker for that belief, and acquaintance with the correspondence that holds between the belief and its truth-maker.

One plausible way out of the problem of the speckled hen for proponents of direct acquaintance is to maintain that in some cases a person may be acquainted with both one's belief that *p* and the fact that *p*, but one could fail to be acquainted with the correspondence that holds between the belief and the fact that *p*.[62] The right move for the defender of acquaintance to answer the problem of speckled hen is to admit that a person can be directly acquainted with the appearance of a hen with 55 speckles and that he can be directly acquainted with the belief that he is having an appearance of a hen with 55 speckles—but that he can fail to be acquainted with the correspondence that holds between the appearance and the belief.

This line of defense can be bolstered by appealing to Richard Feldman's

[62] See Poston (2007) for a similar response to the problem of the speckled hen.

suggestion that there are some properties of our mental experiences that we can fail to grasp because we lack the relevant phenomenal concepts.[63] Phenomenal concepts are the most basic ways in which we think about or recognize conscious experiences. On Feldman's analysis, the difference between most people's ability to have the noninferentially justified belief that one appears to see three red circles and most people's inability to have the noninferentially justified belief that one appears to see a hen with 55 speckles is that most people have a phenomenal concept of appearing-in-a-three-red-circle-way while most people do not have a phenomenal concept of appearing-in-a-fifty-five-speckled-way. In fact, supposing that different people grasp different phenomenal concepts readily explains the difference between typical people who cannot "see" things that idiot savants or those described as having "rainman-like abilities" can "see"—such as the alleged ability to know immediately that one appears to see 148 toothpicks. If one is equipped with the right phenomenal concepts, then one has an important resource for being able to "see" the correspondence that holds between one's beliefs and appearances. Without the right concepts, we are unable to be acquainted with the relevant correspondence relations to have non-inferentially justified beliefs of complex beliefs, such as the belief that one appears to see a 55 speckled hen.

Since I have provided and defended an account of non-inferential justification based on direct acquaintance, the next task is to show how this

[63] Feldman (2004), especially pp. 214-215.

account aids the defender of the Knowledge Argument. Specifically, the next step in my defense of the Knowledge Argument is to substantiate the claim that our knowledge of our qualitative mental states that is secured by direct acquaintance cannot be procured through physical information.

2.3 The Knowledge Intuition and Direct Acquaintance

We clearly know that consciousness has the intrinsic features of experience which we commonly call qualia. As I type on my keyboard and experience the smooth tactile sensations of the keys, I undoubtedly know that my conscious experience has this feature, which is readily accommodated by the epistemology presented in section 2.2. On this account of empirical knowledge, I know that I am experiencing smooth, tactile sensations because I am directly acquainted with that state of mind, the belief that my state of mind has those properties, and the correspondence that holds between the state of mind and the belief. This could be generalized to cover any of the qualitative features of consciousness that we know non-inferentially.

But the knowledge intuition is stronger than merely claiming that one knows that one is having the qualitative features of consciousness. The knowledge intuition also states that our knowledge of these features of consciousness cannot be deduced from physical information alone. Thus, the task for this final section of the chapter is to show how an epistemology based on direct acquaintance substantiates this stronger thesis.

On the account of knowledge defended in section 2.2, there is a reason to think that the qualitative features of consciousness cannot be deduced from physical information alone. The reason is that direct acquaintance epistemically privileges the phenomenal character of consciousness over knowledge of physical information. Recall that my position is a foundationalist one where the basic beliefs cannot be arbitrarily selected (otherwise, the epistemic regress could not be stopped). The non-arbitrary stopping point for this regress is with a belief where one is directly aware that it is a *truth*. In turn, this motivates how the triad of acquaintances can non-arbitrarily provide non-inferential knowledge. The upshot is that our knowledge of qualia does not and (given this way of acquiring knowledge) cannot be deduced from our knowledge of physical truths alone.

2.3.1 Privileging the Phenomenal

In this section, I will argue that phenomenal states of mind are epistemically privileged. Phenomenal states of mind are epistemically privileged because they are that with which we are directly acquainted, and the intrinsic properties of physical objects are not. I am convinced that versions of the arguments from hallucination and illusion demonstrate this. Consider for the sake of illustration the experience of looking out my office window and appearing to see the Iowa River's murky-brown surface. This conscious experience appears the same to me whether it is caused by a veridical experience (e.g., light reflecting from the river that causes my optic nerves to send the right

causal sequence to my brain) or nonveridical experience (e.g., it is caused by a neurosurgeon who is stimulating regions of brain that cause the experience when no river is nearby). In either case my knowledge of my states of mind is privileged compared to my inference as to what is causing my conscious experience. Since in both veridical and nonveridical experiences the conscious experience can be qualitatively identical, the privileged epistemic state is one's knowledge of one's mental state, not the alleged cause of the mental state. After all, people are not directly acquainted with rivers—they are directly acquainted with appearances of rivers or the mental states that exemplify properties that we typically believe to be caused by rivers.

One attempt to deny that veridical and nonveridical experience share a common mental state is disjunctivism. Typically, disjunctivists claim that there is a difference between veridical and nonveridical experiences, but the difference cannot be detected. Given the foundationalist epistemology that requires one to know a truth to stop the epistemic regress, it should be clear why I find this maneuver completely unpersuasive. First, on the basis of direct acquaintance there is no reason to think that even in veridical experiences one has the resources to know noninferentially a truth about the world outside of one's mind. The truth one comes to know noninferentially is a truth about one's state of mind (how it appears), not about the external world. Second, the difference between veridical and nonveridical experiences is typically taken to be a difference in what *caused* the experience. However, this difference is not a

difference that the subject can see from his epistemic position, which essentially makes the salient difference useless from the subject's perspective. To the extent that there is a difference in veridical and nonveridical experience, then, it is worthless from the subject's perspective. But the subject must do his epistemic work with content that is available to his first-person perspective, and from that perspective there is no discernible difference between veridical and nonveridical experiences. Consequently, we find ourselves in the position where we must epistemically prioritize the phenomenal.

The basic form of the classical argument for indirect realism has been articulated in a recent article by Richard Fumerton:[64]

(1) No matter how strong our justification is for believing some proposition describing our immediate physical environment, we could possess precisely the same sort of justification for believing that proposition while vividly hallucinating (while being deceived by a demon, living in the Matrix world, etc.).

(2) The justification we would have were we hallucinating is clearly not noninferential—noninferential justification would require something like our direct acquaintance with some fact about our external environment that is the truth maker for our belief and by hypothesis, there is no such fact.

Therefore,

(3) The justification we have when we are veridically perceiving our physical environment is not noninferential either.

While those who accept alternative accounts of epistemology will resist this argument, it should be clear that given the position I've defended in the previous

[64] Fumerton (2006b), pp. 681-2.

section of this chapter this argument goes hand-in-hand with my other commitments. Thus, from within the camp of a foundationalism based on direct acquaintance, we are forced to accept that there is an epistemic priority of the phenomenal.

2.3.2 Inferential Knowledge of the Physical

In the previous sub-section I presented some reasons for thinking that my account of epistemology implies that there is an epistemic priority for our knowledge of our phenomenal mental states. In this section, I will state what I take to be the best way for this position to characterize the content of our knowledge of the physical. The purpose for providing my account of our knowledge of the physical world is to show why my favored account of epistemology supports the knowledge intuition.

Since I have committed myself to epistemological indirect realism, it follows from my position that one cannot noninferentially know that physical objects exist. One's knowledge of the physical world is mediated by those mental states whose phenomenal content appears to be the same in both veridical and nonveridical experiences. Although my view does not permit one to know directly the nature of physical objects, this should not be taken as an endorsement of skepticism about the external world. Indeed, we have direct access to the effects that physical objects have on us, which may supply ample evidence for acquiring some degree of knowledge about the physical world.

Consequently, our knowledge of the physical world comes primarily from our knowledge of the phenomenal qualities that the physical world causes us to experience. This view should not be confused with a reading of Berkeley's idealism which maintains that we cannot possibly think about mind-independent objects.[65] What I have in mind is more akin to the view of physical objects that Bertrand Russell endorsed in *The Problems of Philosophy*.[66] When we think about physical objects, on my account, we always do so in terms of the causal effects they produce in us. Thus, my understanding of strawberry as a physical object is something like *the cause (whatever it is) that produces the phenomenal experiences of redness, sweetness, firmness, etc.* What we come to know about physical objects, then, is only what we can legitimately infer from the phenomena they cause us to have.

With regard to the knowledge intuition, the upshot for this view of physical knowledge is that we would not expect to be able to infer any (new) knowledge about the phenomenal character of consciousness from physical information alone. In fact, it is the exact opposite. What we know about the physical is epistemically dependent on our knowledge of the phenomenal. I can go even further and claim that if we strip away all of the phenomenal information from the content of our knowledge of physical objects, there would scarce be anything left of our physical knowledge with which to deduce

---

[65] Based on a reading of his so-called "master argument," see Berkeley (1710), secs 23-24.

[66] Russell (1912), especially chapters 2, 3, and 5.

anything.[67] Therefore, given a foundationalist epistemology where all of our knowledge of the physical world is ultimately derived from direct acquaintance, it follows that the knowledge intuition is almost certainly true.

2.4 Concluding Remarks

In this chapter I have made the case for the claim that physical information alone is not sufficient to deduce propositional knowledge that characterizes the qualitative features of our mental states, which I have referred to as the knowledge intuition. In the first part, I presented some thought experiments that support a *prima facie* case for knowledge intuition. Then, after defending a specific approach to empirical knowledge, presented an independent and more substantial defense of the knowledge intuition. The knowledge intuition in conjunction with the physical knowledge intuition that I presented in chapter 1 constitute my positive case for the Knowledge Argument. Even though some may dismiss the commitments that are needed to defend my account of empirical knowledge, they still must find some reason to dismiss the *prima facie* case for the knowledge intuition. In other words, the onus rests on those who claim that the knowledge intuition is false to provide a plausible explanation of its *prima facie* and intuitive plausibility.

---

[67] What content would be left? I imagine we may have some non-phenomenal concepts and logical tautologies that characterize the physical object. However, if Horgan and Tienson (2002) and Pitt (2004) are correct in claiming that all thinking involves a phenomenal qualitative feel, then it may be psychologically impossible for humans to have any thoughts at all that are devoid of qualia.

In the next three chapters I will undertake the task of showing that there are no physicalist alternatives that reasonably defend physicalism against the premises of the Knowledge Argument. Chapter 3 will consider attempts to deny that Mary learns anything new whatsoever when she experiences phenomenal redness for the first time. In the fourth chapter, I will respond to physicalists that accept Mary learns something new, but deny that her knowledge is propositional or factual knowledge. Chapter 5 is a response to physicalists that accept Mary learns new propositional knowledge when she is released from her black-and-white environment but that this is consistent with physicalism. These three chapters makeup my negative case for the Knowledge Argument by showing that there are no acceptable physicalist responses to the argument. It is to these concerns that we turn next.

CHAPTER 3: A STRONG DENIAL OF THE KNOWLEDGE INTUITION, OR DENYING THAT MARY LEARNS ANYTHING NEW

My positive case for the Knowledge Argument has been laid out in chapters one and two. In the first chapter, I defended the Physical Knowledge Intuition (PKI), which states that if knowing all the physical truths is not sufficient for knowing all the truths about the world, then physicalism is false. In chapter two, I supported the Knowledge Intuition (KI), which claims that our knowledge of the subjective character of consciousness cannot be derived from the complete set of physical truths. Those two premises imply that physicalism is false.

Of course, most physicalists have not been content to sit quietly while the Knowledge Argument challenges their position. The next three chapters will explain and rebut the most plausible ways that physicalists have employed to respond to the Knowledge Argument. In this chapter and the next, I will look at two ways to deny the KI. First, I will consider strong denials of the KI, which essentially claim that all knowledge of truths about the world can be derived from all the physical truths. In the following chapter, I will consider weaker denials of the KI, which maintain that all propositional knowledge about the world can be derived from all the physical truths, while recognizing that some non-propositional knowledge cannot be acquired in this way. Chapter 5 will consider physicalist responses that reject the PKI. Once these objections have been dispelled, my case for the Knowledge Argument will be complete.

On the face of it, the strong denial of the KI is wildly counterintuitive. In terms of Frank Jackson's Mary thought experiment, the strong denial of the KI amounts to saying that Mary is able to deduce everything there is to know about the world before she is released and experiences color for the first time. As we saw in the previous chapter, intuitively when Mary experiences color for the first time she comes to know something new, so the strong denial of the KI must explain away this intuition. There are at least two ways that advocates of the strong denial of the KI try to curb the force of this intuition. First, there are those who claim that Mary would in fact be able to figure out what it's like to see red from physical information alone. Instead of being surprised by her first appearance of phenomenal redness, on this view Mary nonchalantly recognizes that her first appearance of phenomenal redness corresponds exactly with the way she figured it would on the basis of the deductions she performed with the physical information alone. A second way for strong deniers of the KI to explain away the intuition that Mary comes to know something new is to show that any apparent knowledge that Mary gains from her first experience of phenomenal redness is actually illusory. If our apparent knowledge of phenomenal redness is illusory, then any alleged knowledge about the phenomenal is not really knowledge about the world at all.

Despite being outrageously counterintuitive on the face of it, strong denials of the KI have some indisputably significant defenders. Probably the two most prominent advocates of this approach include Daniel Dennett and Frank

Jackson (after he defected from his position of being a "qualia freak"). Below I will examine the different ways these two prominent philosophers have argued for the strong denial of the KI.

3.1 Dennett's Strong Denial of the KI

Daniel Dennett tows the hard line against the Knowledge Argument. As he sees it, the argument is a failure because it relies on a bad thought experiment. In his own words, "it is a bad thought experiment, an intuition pump that actually encourages us to misunderstand its premises!"[1] What's wrong with Jackson's Mary thought experiment? Dennett believes that even among the most sophisticated and shrewd thinkers that it is hard to follow directions, especially the part where we are supposed to imagine that Mary possesses *all* the physical information. Since we cannot really imagine this, claims Dennett, most people "just 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."[2] Given that Mary really does possess all the physical information, how are we supposed to know that she cannot figure out what it's like to experience phenomenal redness? Dennett believes that the thought experiment commits a philosophical sleight of hand by giving the false impression that it would be ridiculous to think

---

[1] Dennett (1991), p. 398.

[2] Dennett (1991), p. 399.

Mary could not deduce what it's like to experience phenomenal redness from all the physical information. To make this point, Dennett presents an alternative way to tell Mary's story:

> And so, one day, Mary's captor's decided it was time for her to see colors. As a trick, they prepared a bright blue banana to present as her first color experience ever. Mary took one look at it and said "Hey! You tried to trick me! Bananas are yellow, but this one is blue!" Her captors were dumbfounded. How did she do it? "Simple," she replied. "You have to remember that I know everything—absolutely everything—that could ever be known about the physical causes and effects of color vision. So of course before you brought the banana in, I had already written down, in exquisite detail, exactly what physical impression a yellow object or a blue object (or a green object, etc.) would make on my nervous system. So I already knew exactly what *thoughts* I would have (because, after all, the "mere disposition" to think about this or that is not one of your famous qualia is it?). I was not in the slightest surprised by my experience of blue (what surprised me was that you would try such a second-rate trick on me). I realize it is *hard for you to imagine* that I could know so much about my reactive dispositions that the way blue affected me came as no surprise. Of course it's hard for you to imagine. It's hard for anyone to imagine the consequences of someone knowing absolutely everything physical about anything!"[3]

Dennett explains in a few places that the main idea behind his telling of the Mary story isn't to show that Mary would be able to deduce what it's like to experience phenomenal redness from physical information.[4] Rather, his aim is to show that the original story (and any other similar thought experiment) does not prove that Mary learns something new when she experiences color for the first time. "My variant was intended to bring out the fact that, absent any persuasive argument that this could not be how Mary would respond," writes Dennett, "my

---

[3] Dennett (1991), pp. 399-400.

[4] Dennett (1991), p. 400; Dennett (2005), pp. 105-106; Dennett (2007), p. 16.

telling of the tale had the same status as Jackson's: two little fantasies pulling opposite directions, neither with any demonstrated authority."[5]

Thus, Dennett challenges the intuition that Mary learns something new when she experiences color for the first time. Furthermore, there is a second assumption that Dennett questions:

> Another unargued intuition exploited by the Mary intuition pump comes in different varieties, all descended inauspiciously from Locke and Hume . . . . This is the idea that the 'phenomenality' or 'intrinsic phenomenal character' or 'greater richness'—whatever it is—cannot be constructed or derived by lesser ingredients. Only actual experience (of color, for instance) can lead to the knowledge of what that experience is like.[6]

As this assumption is made explicit, Dennett believes that it threatens to make the Knowledge Argument a trivial exercise.[7] If knowing what it's like to experience phenomenal redness is the same as correctly imagining what it's like to experience phenomenal redness, then imagining what it's like to experience phenomenal redness is nothing more than experiencing phenomenal redness. So, it trivially follows that anyone can't know what it's like to experience phenomenal redness until she has experienced phenomenal redness. In terms of the Mary thought experiment, suppose that physicalism is correct and Mary is able to figure out by use of her deductive and imaginative skills what it's like to experience phenomenal redness (without experiencing it). What should we say in this case? There seem to be two responses. On the one hand, one could claim

---

[5] Dennett (2007), p. 16. Similar remarks appear in Dennett (2005), p. 105.

[6] Dennett (2007), p. 22. Similar remarks appear in Dennett (2005), p. 106.

[7] Dennett (2005), pp. 118-120; Dennett (2007), pp. 23-24.

that she still doesn't know what it's like to experience phenomenal redness (after all, she has only imagined it), but this gives the trivial result that the only way to experience phenomenal redness is to experience phenomenal redness. On the other hand, if one accepts that correctly imagining phenomenal redness is the same thing as experiencing phenomenal redness, then there seems to be no reason (at least one that isn't question-begging) to think Mary can't figure out what it's like to experience phenomenal redness by use of her imaginative and deductive abilities before she is released from her black-and-white environment.

So far I have discussed how Dennett tries to cast doubt on the anti-physicalist intuitions implied by the Mary thought experiment. However, Dennett also offers a positive account as to how Mary would be able to figure out what it's like to experience phenomenal redness. In one place, he suggests a strategy where giving Mary an inch of knowledge about colors, opens the door for her to take a mile:

> [S]he knows precisely which effects—described in neurophysiological terms—each particular color will have on her nervous system. So the only task that remains is for her to figure out a way of identifying those neurophysiological effects "from the inside." You may find you can readily imagine her making *a little* progress on this—for instance, figuring out tricky ways in which she would be able to tell that some color, whatever it is, is *not* yellow, or *not* red. How? By noting some salient and specific reaction that her brain would have only for yellow or only for red. But if you allow her even a little entry into her color space this way, you should conclude that she can leverage her way to complete advance knowledge, because she doesn't just know the *salient* reactions, she knows them all.[8]

[8] Dennett (1991), pp. 400-401.

In his more recent work, Dennett fleshes out this strategy with a new thought experiment, which he takes "to shift the burden of proof."[9] The new thought experiment, involves a robot, which he names RoboMary. Although RoboMary is a robot who has conscious experiences of colors (*which certainly makes the thought experiment more difficult to accept!*), I can let that pass for now in order to see Dennett's point. As I will make clear below, it is possible to show Dennett's error even granting this controversial detail. Dennett provides the RoboMary story in six installments:

1. RoboMary is a standard Mark 19 robot, except that she was brought on line without color vision; her video cameras are black and white, but everything else in her hardware is equipped for color vision, which is standard in the Mark 19.[10]

2. While waiting for a pair of color cameras to replace her black-and-white cameras, RoboMary learns everything she can about color vision of Mark 19s. She even brings colored objects into her prison cell along with normally-sighted Mark 19s and compares their responses—internal and external—to hers.[11]

3. She learns all about the million-shade color-coding system that all Mark 19s have.[12]

4. Using her vast knowledge, she writes some code that enables her to colorize the input from her black-and-white cameras (à la Ted Turner's cable network) according to voluminous data she gathers about what colors things in the world are, and how Mark 19s normally encode these. So now when she looks with her black-and-white cameras at a ripe banana, she can first see it in black and white, as pale gray, and then imagine it as yellow (or any other color) by just engaging her

[9] Dennett (2007), p. 30.

[10] Dennett (2007), p. 25. Similar remarks appear in Dennett (2005), p. 122.

[11] Dennett (2007), p. 26. Similar remarks appear in Dennett (2005), p. 123.

[12] Dennett (2007), p. 26. Similar remarks appear in Dennett (2005), p. 123.

colorizing prosthesis, which can swiftly look up the standard ripe-banana color-number profile and digitally insert it in each frame in all the right pixels. After a while, she decides to leave the prosthesis turned on all the time, automatically imaging the colors of things as they come into focus in her black-and-white camera eyes.[13]

5. She wonders if the ersatz coloring scheme she's installed in herself is high fidelity. So during her research and development phase, she checks the numbers in her registers (the registers that transiently store the information about the colors of the things in front of her cameras) with the numbers in the same registers of other Mark 19s looking at the same objects with their color-camera eyes, and makes adjustments when necessary, gradually building up a good version of normal Mark 19 color vision.[14]

6. The big day arrives. When she finally gets her color cameras installed, and disables her colorizing software, and opens her eyes, she notices . . . nothing. In fact, she has to check to make sure she has the color cameras installed. She has learned nothing. She already knew exactly what it would be like for her to see colors just the way other Mark 19s do.[15]

Dennett is aware that many will find the fourth installment of the RoboMary story to be question begging or cheating of some other sort.[16] So, he tells a version of the story where RoboMary is "locked" or prohibited from altering her own visual system. He contends that RoboMary could still figure out what it's like to experience phenomenal redness, even if the system is locked. RoboMary could build an exact duplicate of herself that has the capacity to experience colors, and based on her extensive knowledge of robots and the physics of color, she could determine how the replica of herself would react in

[13] Dennett (2007) pp. 26-27. Similar remarks appear in Dennett (2005), p. 124.

[14] Dennett (2007), p. 27. Similar remarks appear in Dennett (2005), pp. 124-125.

[15] Dennett (2007), p. 27. Similar remarks appear in Dennett (2005), p. 125.

[16] Dennett (2005), pp. 126-128; Dennett (2007), pp. 27-30.

every possible color situation. RoboMary could observe her dispositional state when she looks at a red tomato with her black-and-white system and compare it with the dispositional state of her duplicate that has color-viewing capacities. Without altering her visual systems, then, RoboMary could make all of the adjustments to put herself in exactly the same dispositional state of her color-viewing counterpart when it is looking at a ripe tomato.[17] Dennett concludes that "now she can know just what it is like for her to see a red tomato, because she has managed to put herself into just such a dispositional state. . . ."[18]

We can sum up Dennett's criticisms against the Knowledge Argument with three points. First, he claims that the thought experiment misleads the reader to believe the erroneous intuition that Mary comes to learn something new when she experiences color for the first time. Second, he questions the assumption that the phenomenal character of experience cannot be known through lesser ingredients. Third, given Mary's immaculate knowledge of the internal and external effects of colors on her subjects, he believes that one can more plausibly understand how Mary comes to know what it's like to experience phenomenal redness from physical information alone (as illustrated through

---

[17] This is similar to the possibility of "Swamp Mary" – another thought experiment where a creature, Swamp Mary, emerges from a freak accident where her brain is in the same state as a person who has experienced phenomenal redness in the past and remembers what it is like, although Swamp Mary never has experienced phenomenal redness. It would seem that Swamp Mary knows what it is like to experience phenomenal redness even though she never has actually experienced it for herself. See McGreer (2003); Dennett (2005), pp. 120-122; Dennett (2007), p. 24, pp. 28-29. Compare with van Gulick's example of DRamy in his (2004), p. 386.

[18] Dennett (2007), p. 28. The rest of the sentence compares this result with Swamp Mary. See footnote 17 of this chapter, if interested in knowing more about Swamp Mary. Similar remarks appear in Dennett (2005), p. 128.

RoboMary). I will examine each of these facets of Dennett's criticisms in turn below.

In response to Dennett's first point that the thought experiments used in the Knowledge Argument mislead the reader to accept a dubious intuition, I have two rejoinders. First, in my defense of the Knowledge Argument I have provided more than raw intuition ginned up by thought experiments to underwrite the knowledge intuition. In sections 2.2 and 2.3, I presented and defended an approach to epistemology that explains why the knowledge intuition is correct. So, my defense of the Knowledge Argument does not rely solely on the impact of the thought experiments to convince readers that the knowledge intuition is correct. Nothing Dennett has written directly addresses the possibility that the knowledge intuition is secured by a defensible approach to epistemology.

Second, Dennett's attempt to cast doubt on the knowledge intuition is bolstered entirely by his alternative thought experiment with the blue banana trick. Recall that the point of his alternative thought experiment is that it had equal viability as Jackson's. As he explained, "my telling of the tale had the same status as Jackson's: two little fantasies pulling opposite directions, neither with any demonstrated authority."[19] The problem, however, is that Dennett's alternative thought experiment with Mary is not equally persuasive as Jackson's original. If it can be shown that Jackson's thought experiment is more appealing

[19] See footnote 5 above for references.

than Dennett's, then Dennett's attempt to undercut Jackson's original thought experiment will fail to carry any force.

The operative question, then, is why should we think that Dennett's telling of the Mary story is less plausible than Jackson's? One initial problem is that it is difficult to see how Dennett's story is supposed to stand in opposition to Jackson's.[20] In Dennett's version of the story, he highlights that Mary wouldn't be fooled by the blue banana trick, but his story doesn't reveal how Mary comes to know what the subjective character of a color experience is like prior to her release. Rather, all Dennett's story shows is that Mary is able to distinguish between some of the natural colors, which is something that Mary could do without relying on the color-qualia experiences (for example, by noticing the physiological and behavioral reactions in herself to the colored object). We could even imagine that Mary wouldn't be fooled by the blue banana trick even in her black-and-white laboratory. The point of the Knowledge Argument, of course, is that Mary learns something new about the intrinsic character of color qualia when she leaves the lab. If Dennett's banana trick doesn't reveal that Mary knows the intrinsic character of phenomenal yellowness prior to her release, then it cannot offset the KI presented in Jackson's argument. In other words, the point of Jackson's thought experiment did not focus on Mary's clever ways of identifying colors; rather, the point is that Mary didn't know what the phenomenal character of color experience was like. Dennett's thought

[20] Similar criticisms appear in Jacquette (1995), pp. 226-227 and Alter (1998), pp. 44-45.

experiment at best shows Mary's cunning ability to identify colors, but it does not show that it is plausible to believe Mary comes to know the phenomenal character of color experience.

But even if we reinterpret Dennett's thought experiment so that it is about the phenomenal character of conscious experience, there is still a reason to doubt that it is not as persuasive as Jackson's telling of the Mary story. On this reading, Dennett's story attributes to Mary the knowledge of the intrinsic features of color experience somehow from her knowledge of the physical properties of colored objects and the dispositions of human psychology toward those properties. The problem is that Dennett is open to the charge of confusing knowledge of two different sorts of things.[21] There is a difference between possessing knowledge that enables her to pass a behavioral test involving color identification and knowing what it's like to experience phenomenal redness—but Dennett is running these together. In fact, Dennett explicitly claims that this difference (which most people readily grasp) does not exist. In a footnote discussing the distinction between knowing *what one would say and how one would behave* and knowing *what it's like*, he writes, "If there is such a distinction, it has not yet been articulated and defended." He continues, "If Mary knows *everything* about what she would say and how she would react, it is far from clear that she wouldn't know what it would be like."[22] Thus, in order to accept Dennett's alternative

[21] See Robinson (1993a) and Jacquette (1995).

[22] Dennett (2006), p. 106, n. 3.

account of Mary as equally plausible as Jackson's it would require accepting something incredible—namely, that there is no distinction between a person's knowing the dispositions to behave to certain stimuli and a person's knowing what it's like to undergo certain phenomenal experiences. Therefore, Dennett's account scores far lower than Jackson's on the scale of *prima facie* acceptability.[23]

So much for trying to gauge initial intuitions—now let's consider whether Dennett's last two criticisms of the Knowledge Argument can be sustained. Does Dennett make good on his claims that the phenomenal character of experience can be known through lesser ingredients and that RoboMary presents a plausible way to see how one can come to know the phenomenal character of experience through non-phenomenal constituents? Since these two objections are ideologically linked, I will respond to them together.

First, I will shamelessly note that my defense of the Knowledge Argument presents a straightforward way to respond to Dennett's last two points. The phenomenal character of experience cannot be known through its non-phenomenal components on my view because of the epistemic priority of the phenomenal (see §2.3). Thus, on my account the intuitive ideas that undergird the Knowledge Argument are not unargued hunches, but they are corollaries of an epistemological framework.

---

[23] The *prima facie* plausibility of the distinction, if proof is necessary to support it, is evident from the well-known philosophical discussion on inverted qualia, absent qualia, and dancing qualia. See especially Block (1978), Block (1990), Chalmers (1995b), and Chalmers (1996), pp. 247-275. See Nida-Rümelin (1996) for empirical grounds for taking qualia inversion seriously. For an overview with bibliography see Byrne (2006).

But even apart from my controversial account of epistemology, there are good reasons to doubt Dennett's claim that the phenomenal character of experience can be known through its non-phenomenal parts. The very notion that the phenomenal character of consciousness can be known through non-phenomenal information is quite incredible, so I will lay the onus on Dennett to convince anyone otherwise. Thus, the strength of Dennett's objection rests on the plausibility of his RoboMary thought experiment.

In his first pass of the RoboMary story, Dennett shows how an unlocked RoboMary could reconfigure her visual systems to grant her the same color information as her color-endowed counterparts. But as Dennett suspects, the unlocked version cheats. The cheat occurs by allowing RoboMary to induce the state of seeing color by changing her visual systems. One reason to call this cheating is that it equates RoboMary's altering of her visual systems with the abilities of human imagination, which is not comparable. Michael Beaton explains, "there is no reason to think that we humans have the ability to configure our low level colour processing circuitry the way unlocked RoboMary does, just by thinking about it, in advance of any exposure to colour."[24] Another reason to think that this is cheating is that it puts RoboMary in the state of seeing color without deducing what that state is like from her physical information. The point of the Knowledge Argument is to show that our knowledge of what conscious experience is like cannot be deduced from physical information alone.

[24] Beaton (2005), p. 18.

Dennett shows that RoboMary can know the cause of phenomenal states and she can implement those causes in herself, and thereby she can come to know what it's like to experience color. So the unlocked version of RoboMary does not show that phenomenal knowledge can be deduced from non-phenomenal parts.

This means that Dennett's case against the Knowledge Argument rests entirely on the plausibility of locked RoboMary to figure out what it's like to undergo conscious experience. Contrary to what Dennett claims, however, locked RoboMary doesn't come to know what it's like to see color. Locked RoboMary only comes to know how to mimic the verbal and external behavioral dispositions of one of her counterparts with a color-enhanced visual system. But as I noted above, there is a difference between knowing the external dispositional behavior that typically accompanies phenomenal knowledge and knowing the phenomenal character of conscious experience. Dennett, of course, thinks that there is no difference between knowing these apparently different things. In response to an objection based on RoboMary's inability to use phenomenal concepts to pick out thoughts demonstratively, Dennett defends his point in his characteristic style:

> Why can't RoboMary form demonstratives that allude to the relevant states of her model, instead of her own locked color system? And why wouldn't they be just as good? Because they wouldn't have that extra *je na sais quoi*? But that is just what has not been shown to exist. In the case of RoboMary, the temptation to posit a rather magical extra property that adheres somehow to her entering into these color-system states (which are basically just numbers in registers, after all) is weak. The temptation should be resisted in the case of Mary, too. It has no legitimate business to

do and tends to distort the imagination covertly.[25]

For Dennett's objection to succeed, it seems that everything depends on whether there is something more to knowing what it's like to experience phenomenal redness than knowing the dispositional states that a person is in when he is experiencing phenomenal redness. Since Mary and RoboMary can plausibly come to know the dispositional states of people who see the color red, then, the crucial question is whether that is the same thing as what it's like to experience phenomenal redness. Despite the fact that I believe most people share the strong intuition that there is something more to experiencing redness than having the right verbal and behavioral dispositional states, nonetheless I will offer three reasons to think that there is a significant difference.

First, it is logically possible that a person can have the external dispositional behavioral states of experiencing phenomenal redness and yet not know the phenomenal character of redness. Here the inverted spectrum thought experiments illustrate the point.[26] It is logically possible that tomorrow Brett wakes up and discovers that all of his color experiences are "inverted"—that is to say that objects that typically cause normal-sighted people to see them as green now appear to Brett as though they are red (and vice versa). Given a complete, systematic inversion of his color experiences, there is no contradiction in supposing that Brett would (given enough time) behave with the exact same

[25] Dennett (2007), p. 29.

[26] See note 23 above for references.

dispositions as normal-sighted people. In other words, Brett exemplifies all the external dispositional states of seeing red when a red object is put in front of him, but the object appears phenomenally green to him. The logical possibility of the inverted spectrum thought experiment shows that there is a difference between having the external dispositions to behave towards colored objects and knowing the intrinsic character of phenomenal experiences of color. Given this difference, Dennett's response to the Knowledge Argument fails because it presumes that there is no difference.

Those familiar with Dennett know that he is skeptical of thought experiments and what implications can be drawn about the actual world from what is logically possible. So, I will offer a more empirical reason to deny that having the external dispositions to behave towards a color is the same as knowing the phenomenal character of a color. There is a well-documented case of blindsight, which shows that there are actual cases of people who exhibit the external dispositional behaviors towards information typically acquired by vision, but who lack the phenomenal character of visual experience.[27] For example, with type 1 blindsight the subject has no awareness whatsoever of any visual stimuli, yet the subject can still accurately report information that is normally acquired through vision, such as the location or movement of an object, far beyond the statistics of chance. Given that the empirical research that supports the existence of blindsight is good, the distinction between a person's

[27] For example, Weiskrantz (2009).

disposition to behave towards this information and the intrinsic character of visual experience is a distinction that is exemplified in the actual world.

My third reason comes from another actual case where the difference between being disposed to behave towards visual stimuli appropriately and knowing the intrinsic qualities of visual experience is conveyed through research in Tactile Vision Substitute Systems (TVSS).[28] TVSS takes visual information gathered through a video camera and transduces the information in a tactile way (such as through vibrations or electric charges), which can be received by a blind person's skin. After proper training, blind people equipped with TVSS have been able to make accurate judgments about the information transduced by the camera (describe shape, location, and movement of objects) and perform impressive tasks that involve hand-eye coordination such as batting a ball, working on an assembly line, and riding a bicycle around obstacles. Despite demonstrating amazing consistency to respond with the right behavioral dispositions toward this information, people with TVSS fail to report any phenomenal qualia. Paul Bach-y-Rita relays the following results from his research:

> Subjects trained with the tactile vision substitution system have noted the absence of qualia, which in a number of cases has been quite disturbing. Thus, well-trained subjects are deeply disappointed when they explore the face of a wife or girlfriend and discover that, although they can describe details, there is no emotional content to the image. In two cases, blind university students were presented *Playboy* centerfolds, and although they could describe details of the undressed women, it did not have the

[28] One place to find information on TVSS is Bach-y-Rita (1996).

> affective component that they knew (from conversations with their sighted classmates) that it had for sighted persons.[29]

Once again, empirical studies of TVSS demonstrate that the distinction which Dennett rejects is not merely a conceptual possibility. The difference between having the right behavioral dispositions and knowing the intrinsic character of phenomenal visual experience is a real distinction that can be seen in cases from the actual world. Therefore, Dennett's obstinate stance against this distinction cannot be maintained in the face of our actual empirical studies.

Perhaps those who take the having of dispositions to behave towards color to be the same as knowing what it's like to see color will insist that all three examples given above fail to confer *all* the behavioral dispositions of seeing color, which explains why people can have those dispositions and yet fail to know what it's like to see color. At this point, however, the position seems to be a very implausible sort of dogmatism about the nature of phenomenal knowledge. The examples given above present conceptual and empirical reasons to think that having behavioral dispositions cannot be the same thing as knowing the intrinsic character of color experience. To insist that adding more behavioral dispositions to these examples would enable the subjects in them to know the phenomenal character of color experience does not appear to be remotely plausible to me. A much more plausible explanation is to admit that in each of the cases noted above adding more external dispositions to the subjects will not dispel their

---

[29] Bach-y-Rita (1996), p. 509.

ignorance of the phenomenal character of experiencing a certain color.

The final problem with Dennett's locked RoboMary example is that it fails, like the unlocked RoboMary case, to demonstrate that RoboMary is able to know what it's like to see red by deducing it from the physical information. Locked RoboMary doesn't deduce what it's like to see red from the physical information. Rather, she performs an experiment on her counterpart and physically puts herself into a similar state. Instead of meeting the challenge presented by the Knowledge Argument, which questions whether knowing all physical truths is sufficient to know all the truths about the world, Dennett sidesteps the issue altogether and contents himself with RoboMary's putting herself in what he takes to be the relevant state of knowledge. Once again, this appears to be a form of cheating. "I just don't see that this is what matters. So far as I can see," Dennett responds,

> this objection presupposes an improbable and extravagant distinction between (pure?) deduction and other varieties of knowledgeable self-enlightenment. I didn't describe RoboMary as "programming" herself; I said she "notes all the differences between state A, the state she was thrown into by her locked color system, and state B, the state she would have been thrown into had her color system not been locked, and—being such a clever, indefatigable, and nearly omniscient being—makes all the necessary adjustments and *puts herself into state B*." If I use my knowledge to imagine myself into your epistemic shoes in some regard, is this "self-programming"? And if so, what is the import of this characterization for the knowledge argument?[30]

In short, either RoboMary can deduce the knowledge of what it's like to see red from the physical information or she cannot. Dennett admits that she

[30] Dennett (2007), p. 29.

cannot. Instead, Dennett explains that RoboMary can do something else to figure out what it's like to see red; namely, she can figure out the state that causes one to experience what it's like to see red and then she can put herself in that state. Thus, the sense in which RoboMary "figures out" what it's like to see red does not involve deductive inferences (or even inductive inferences) from the physical information to the content of the experience of seeing red. The problem, then, is that Dennett casually equates "powers of imagination" with actually putting oneself in the state which causes certain experiences. But this is not at all the same as using one's imagination.

At this point, Dennett may complain that this trivializes the Knowledge Argument. If Mary cheats by putting herself in the state that causes her to experience phenomenal redness and the only way to know what it's like to experience phenomenal redness is to be in that state, then any physicalist attempt to show how Mary could come to know what it's like to experience phenomenal redness will count as cheating. If this point is correct, Dennett claims, "then we philosophers have been wasting a lot of time and energy on what appears to be a relatively trivial definitional issue."[31] Perhaps the point is elementary, but it is hardly trivial. It is a remarkable insight to realize that the truths that correspond to the phenomenal character of conscious experience are not capable of being known through the content of physical information alone. Contrary to Dennett, then, this does not make physicalism false in a trivial way. Since the phenomenal

[31] Dennett (2007), p. 24. Similar remarks appear in Dennett (2005), pp. 119-120.

features of conscious experience could have failed to exist, physicalism could have been true. Furthermore, physicalists can find other ways to deny that Mary learns anything new when she leaves the black-and-white lab (such as Jackson's latest position), or physicalists can acknowledge that Mary learns something new which is compatible with the tenets of physicalism (to be discussed in chapters 4 and 5). Given these alternatives, Dennett's charge that the Knowledge Argument appears to be a "relatively trivial definitional issue" rings hollow.

I conclude, then, that Dennett's case for the strong denial of the knowledge intuition fails. His thought experiments fail to have equal probative force as those that support the Knowledge Argument, and much of his positive account equivocates on important concepts or smuggles in Mary's new knowledge through a cheat. However, this does not mean that the physicalist strategy that strongly denies the knowledge intuition is hopeless. Dennett denies that Mary learns anything new when she experiences color because he plays up what Mary could figure out in the black-and-white lab. However, another way to deny that Mary learns anything new is to downplay the new information that she gains when she is released. This second strategy is considered by Frank Jackson, which I shall assess next.

3.2 Jackson's Strong Denial of the KI

Frank Jackson once was the foremost promoter of the Knowledge Argument against physicalism. It is well-known that Jackson changed his mind

in the mid-1990s, and now he thinks that physicalism is not threatened by the Knowledge Argument. How could the man who penned the famous Mary thought experiment come to reject the argument he once thought was so convincing? The answer is found in Jackson's embracing a theory of representationalism. Before discussing what representationalism is and how Jackson takes it to falsify the KI, I briefly need to make two points. First, Jackson remains an indefatigable defender of *a priori* physicalism, and therefore he does not doubt the PKI. Second, Jackson's official position is that representationalism and the ability hypothesis taken together undermine the knowledge intuition. Since I won't be discussing the ability hypothesis until the next chapter, the substance of this review of Jackson's position surveys the strength of representationalism to cast doubt on the KI.[32] Of course, since I will be arguing in the next chapter that the ability hypothesis does not sufficiently explain away the knowledge intuition, perhaps readers should wait until reading that chapter to decide whether Jackson's change of mind has been for the better.

Representationalist theories of perception state that we perceive the world through the objects represented in conscious experience. In typical cases of perception, the representationalist will say that perceivers are aware of objects and their represented properties in their experience, but that they are not aware

[32] However, Jackson (2002), p. 439, claims that he cannot see how the ability hypothesis could be correct if it wasn't for accepting his form of representationalism. This could reasonably be taken to imply that if representationalism fails in its task to answer the Knowledge Argument, then Jackson doesn't think the ability hypothesis is a plausible response to the argument.

of the intrinsic character of experience itself. On Jackson's representationalism, "accessing the nature of the experience itself is nothing other than accessing the properties of its object."[33] This is because in typical cases, the nature of experience itself is transparent or diaphanous.[34] Furthermore, on the strong version of representationalism that Jackson holds, the content of experience is exhaustively representational. The motivation for accepting the exhaustive thesis is that if experience consisted of both representational and non-representational components, it should be possible to vary the non-representational part while leaving the representational part the same. However, it is not possible to do so. Once you change the kind of experience one is having, it changes how things are represented to be (or so the argument goes). Thus, the argument for strong representationalism must hold that there is no possible way of altering a conscious experience without altering the representational content of the experience.[35]

The representationalist theory of perception is sometimes called an intentionalist theory of perception because it requires all perception to take an object. In other words, perception is always *about* something; when there is perception, then there is always something that is being perceived.

Physicalists who wish to endorse representationalism to avoid the

---

[33] Jackson (2007b), p. 55.

[34] Other representationalists with similar views as Jackson's are Harman (1990) and Tye (2000).

[35] For his argument for strong representationalism, see Jackson (2007b).

consequences of the Knowledge Argument clearly need to disavow the notion that the objects of perception are always real objects. Classical sense datum theorists such as G. E. Moore have endorsed a kind of representationalism, but they emphasize that the objects of perception are real objects, which is why they take sense data to be real.[36] Since reifying sense data is not compatible with physicalism, the representationalist must take a different route. The standard line is for representationalists to maintain that the objects of perceptual experience are *intensional objects*. What is an intensional object? Jackson describes his position on the nature of an intensional object this way:

> 'it' is not an object at all, and our use of verbal constructions that belong to the syntactic category of names is a convenient, if metaphysically misleading, way of talking about how things are being represented to be. We talk of being directly aware of a square shape in our visual fields, but there is no square shape to which we stand in the relation of direct awareness; rather, our visual experience represents that there is something square before us. What makes it right to use the word 'square' in describing our experience is not a relation to something that has the property the word stands for but the fact that the way the experience represents things as being can be correct only if there is something square in existence. Thus on this view the squareness of an experience is an intensional property, not an instantiated one. The same goes mutatis mutandis for all the properties we ascribe to what is presented in experience, the properties we have in mind when we talk of the properties we ascribe to what is presented in experience, the properties we have in mind when we talk of the qualities of experience and to which the argument from diaphanousness applies. When we use words like 'square', 'two feet away', and 'red' to characterize our experiences, we pick out intensional properties, not instantiated ones.[37]

On Jackson's representationalism, then, the mind represents the world as

---

[36] Moore (1922).

[37] Jackson (2002), 427-428.

being a certain way, but there is no commitment that anything is actually the way the mind represents it to be. For example, there can be misrepresentations, such as when the mind represents a straight stick submerged in water as being bent. Just because the object of perception has the property of being bent, representationalists do not concede that there must be something that has the property of being bent. The representationalist of Jackson's kin maintains that in non-veridical representations (such as the submerged straight stick), there is nothing that has the misrepresented property (such as the property of being bent). Since the intensional "object" itself is not a real object at all, it does not instantiate the misrepresented property.

Representationalism appears to be a useful tool for denying the KI. If the contents of the mind consist of intensional objects and some intensional objects are misrepresentations (e.g., there is no real object with the property that is being represented), then one way to diagnose what happens when Mary leaves the black-and-white environment and sees a red tomato is to accept that the redness of her mental representation is illusory. Although Mary has a mental representation of phenomenal redness, it doesn't follow that anything real has the property of being phenomenally red. Jackson explicitly describes his view as taking color experience as illusory: "there is a pervasive illusion that conspires to lead us astray when we think about what it is like to have a color experience."[38]

Jackson provides both a strong and a weak thesis to capture the illusory

[38] Jackson (2002), p. 422.

nature of color experience. The strong thesis is that "we should be eliminativists about red and about color in general."[39] This would imply that nothing is red (or colored), and that any representation of the world being red (or colored) is a misrepresentation of the way things are. There is a weaker position that Jackson describes where "our experience of color contains a substantial degree of misrepresentation—the misrepresentation that leads dualists astray—[but] there are complex physical properties 'out there' that stand near enough to those captured by the color solid for us to be able to identify them with the various colors."[40] On the weaker thesis, then, nothing is phenomenally red (or colored), but physical objects may have complex physical properties that are very similar to the colored properties of our experiences, and thus they only slightly misrepresent how the world is. Presumably, these physical properties that are similar to the phenomenal properties of color experience are the sort of properties that Mary would be able to know in the black-and-white lab. Jackson does not take a position on whether he endorses the strong or the weak thesis. In either case, the property of phenomenal redness is illusory, and the only truths that exist are those that can be known from Mary's black-and-white lab. Given Jackson's particular brand of representationalism, it follows that the KI is false. When Mary sees a red tomato for the first time, according to Jackson, she learns no new truth about the way the world is.

---

[39] Jackson (2002), p. 432.

[40] Jackson (2002), p. 432.

My criticisms of Jackson's strong denial of the KI will focus on three points. First, I will make the case that even if representationalism answers the KI, the physicalist who desires to keep his metaphysics austere should have no motivation to accept the metaphysical commitments of representationalism. Second, I will argue that representationalism by itself does not have the resources to mount a reasonable case against the KI. Finally, I will argue that Jackson's representationalism is false, or at least that it is implausible.

The key move in the representationalist's denial of the KI is to allow that not all mental representations veridically represent the way the world is. In other words, by allowing the objects of representation to be intensional objects, the representationalist is not committed to saying that whatever is represented must exist. So, when Mary leaves her black-and-white surroundings and starts having mental representations of red things, it is open to the representationalist to claim that Mary is misrepresenting the way the world is. Even though Mary is now having representations of the world as being red, the representationalist could try to save physicalism by claiming that these representations are actually massive, systematic misrepresentations. But the physicalist's salvation comes at an extravagant price. By embracing intensional objects as the "objects" of representation, the physicalist is now essentially embracing the Meinongian category of non-existent objects.

Recall that representationalism takes all conscious experience to be intentional. Thus, all conscious experiences are about something. Sometimes

conscious experience is about something real, such as when one is having representations of the world being the way it actually is. Other times, however, conscious experience is about something unreal, such as when one is having misrepresentations about the way the world is. Since all conscious experience is intentional, according to representationalists, they cannot deny that there is an object that the experience is about. As a type of physicalism, representationalism cannot follow the traditional sense datum theory and reify the properties of conscious experience.[41] This leaves the physicalist who is wielding representationalism in the uncomfortable position of admitting a plethora of unreal objects of conscious experience.[42] If accepting non-physical properties leave physicalists with a bad taste in their mouths, then why doesn't the category of unreal objects or unreal properties seem equally unappealing? As far as I can tell, the physicalist has no motivation for accepting the metaphysical baggage of representationalism—especially since avoiding metaphysical baggage is typically what motivates physicalism in the first place. It seems to me that accepting unreal objects or properties is far more problematic on the face of it than accepting non-physical properties. In any case, the physicalist should be nervous if the only way to do away with non-physical properties is to embrace unreal objects. My point here is that representationalism seems to have no motivation

---

[41] This would obviously play right into the dualist's Knowledge Argument.

[42] In addition to Jackson's account of intensional objects (that corresponds with note 37 above) see also Tye (2000), pp. 47-48, 109-111; and Lycan (2008), §4.2 for examples of other representationalists who accept unreal objects as part of representationalism.

or ontologically satisfying alternative for the physicalist who is concerned about increasing his ontological commitments.

The second problem with the physicalist using representationalism to reject the KI is that it cannot be used to mount a reasonable case against the KI by itself. Even if one grants that it is *logically possible* to reject the KI via representationalism, this isn't going to suffice to show that it is *reasonable* to reject the KI via representationalism. One of the chief challenges in Torin Alter's response to Jackson's latest attempt to evade the Knowledge Argument is that there is no reason why a representationalist of Jackson's sort cannot accept that the represented properties of conscious experience are non-physical properties.[43] What reasons are there to suppose that one's conscious experience of phenomenal redness is a systematic, illusory misrepresentation of the world? In answering this question, Jackson may be guilty of a physicalist sort of "tub-thumping," which he accuses dualists of committing when they employ the Knowledge Argument:

> Intensionalism means that no amount of tub-thumping assertion by dualists (including myself in the past) that the *redness* of seeing red cannot be accommodated in the austere physicalist picture carries any weight. That striking feature is a feature of how things are being represented to be, and if, as claimed by the tub thumpers, it is transparently a feature that has no place in the physicalist picture, what follows is that physicalists should deny that anything has that striking feature.[44]

Here, Jackson contends that the reason one should deny that phenomenal

---

[43] Alter (2007).

[44] Jackson (2002), p.431.

redness is a property of the actual world is that it is inconsistent with physicalism. If this is supposed to be an argument for taking the alleged phenomenal properties of consciousness that satisfy the knowledge intuition to be systematic misrepresentations, it is hardly a non-question-begging reason for saving physicalism from those alleged properties! In the explanation given in Jackson's quoted passage above, the reason for counting qualia as misrepresentations of reality is that they are inconsistent with physicalism. If "tub-thumping" amounts to begging the question, then Jackson is guilty of tub-thumping in his argument for counting qualia among the unreal properties of mental representation.

But is the dualist or the representationalist-physicalist being more incredulous here? Thus far I am granting that representationalism may be true. But this does not by itself inform us how to tell which intensional objects represent the world (and thereby are real) or fail to represent the world (and thereby are unreal). As far as I can tell there is no legitimate reason to take the alleged properties of qualia to be illusory misrepresentations, unless one has already decided prior to considering the Knowledge Argument that all of our philosophical theories must accommodate physicalism. If the physicalist is merely appealing to representationalism to show that given the apparent features of conscious experience physicalism could possibly (in the broadest sense of "possibly") be true, then the dualist may concede the point. What the physicalist still needs to show, however, is that representationalism makes it

reasonable to believe that the purported phenomenal properties of conscious experience are systematic misrepresentations of the actual world, which is to say that they are unreal. Pointing out that these representations *could be* unreal is not the same as showing that they must be unreal or they are most reasonably believed to be unreal.

Perhaps, one possible route for the physicalist to argue that it would be reasonable to believe phenomenal properties misrepresent reality is by appealing to the claim that the content of conscious experience is exhaustively representational.[45] In terms of representationalism, the dualist appeals to the property of the intensional object to underwrite one's knowledge that a property like phenomenal redness exists. But the phenomenal character of experience, according to Jackson's strong representationalism, is exhausted by its representational content. Thus, when Mary sees a red tomato for the first time, she has the state of representing that *something red and round is before me*. Since Mary can know the representational content of this state prior to her release, there is no need to think there is anything missing from her knowledge of the world prior to her release. (After all, the representation of something's being red does not necessarily require the mode of representation to be phenomenally red.) Thus, the phenomenal character of consciousness is discarded with the intensional object used to represent the world.

The best way to argue against using strong representationalism to write

[45] This argument is explicitly presented in Jackson (2007b).

off the reality of phenomenal properties of conscious experience, I believe, is to show that strong representationalism is very implausible. In particular, I will focus my attention on the strong representationalist's claim that the phenomenal character of conscious experience is exhausted by its representational content. If this claim is false, then representationalism in and of itself cannot aid the physicalist in denying the knowledge intuition.

My contention, then, is that it is not the case that the phenomenal character of conscious experience is exhausted by its representational content. In other words, if strong representationalism is true, then there can be no difference in phenomenal content without there being any difference in mental representation. Additionally, if the phenomenal character of conscious experience is exhausted by its representational content, then it would not be possible for there to be qualia that do not represent anything. But some qualia are not intentional, that is to say that they do not represent anything. That constitutes a second reason to reject strong representationalism. Third, I will argue that strong representationalism is false because it is possible to differ the phenomenal content while having the same representational content. If strong representationalism is correct, then the phenomenal content of an experience is exhausted by its representational content. But if it is possible to have the same representational content with different phenomenal content, then there must be something more to phenomenal content than its representational content. Let's take each of these objections in turn below.

First, let's consider the possibility that there can be a difference in phenomenal content without a difference in mental representation. Ned Block's example of inverted earth demonstrates this sort of counterexample to strong representationalism.[46] On inverted earth, the inhabitants' speech resembles English on normal earth, except that the intentional content of their color concepts are inverted with respected to ours. When inhabitants on inverted earth say "red," they mean green. Furthermore, the real physical colors of objects are somehow perfectly inverted as well. Now if you take a person from normal earth (let's call him "Norm") and invert his color vision and move him to inverted earth, then Norm will have the same phenomenal experiences he had on normal earth but the intentional content will change. For example, when Norm looks at a stop sign on inverted earth, it will have the same phenomenal redness that accompanies stop signs on normal earth. Yet, the intentional content of Norm's mental representation of the stop sign on inverted earth refers to the property of something's being green. Hence, it is possible to have the same phenomenal content and different intentional or representational content. Therefore, strong representationalism is false.

In response to Block's inverted earth counterexample, many representationalists have maintained that the phenomenal content of someone who undergoes Norm's inversions would actually change.[47] In other words, two

---

[46] Block (1990).

[47] Such as Dretske (1995), Dretske (1996), Lycan (1996), Lycan (2001), and Tye (1995).

people who are in brain states that are molecule-for-molecule identical could have different phenomenal experiences. Following Dretske's terminology, I will refer to this defense as "phenomenal externalism."[48] The rough idea is that the phenomenal character of conscious experience is determined by its representational content and the representational content includes information that is not constituted entirely by the internal states the subject who is having that conscious experience. If phenomenal externalism is true, then the inverted earth example fails to show that it is possible to have the same phenomenal content and different representational content.

Phenomenal externalism, however, is deeply counterintuitive. It is difficult to motivate the position that the complete content of conscious experience is not constituted by the internal mental states of the conscious subject. I cannot help but think that those who argue for such a transparently false position must mean something different when they use terms like phenomenal content or conscious experience. For example, when phenomenal externalists say that two people with indistinguishable internal mental states can have different phenomenal content due to external factors, they surely don't mean that the "mental twins" have different phenomenal experiences. Two people in the exact same mental state would have no difference in conscious experience, no matter how one spells out the external factors. Phenomenal externalism is an untenable solution, then, because it either trades on an

[48] Dretske (1996).

ambiguity with regard to phenomenal content to save representationalism from the inverted earth counterexample or it is obviously false by denying that conscious experience of mental twins could be different by altering external factors.

My second problem with representationalism is that there are some qualia that are not intentional, which implies that some qualia are not exhausted by their representational content.[49] Perhaps you've had the experience of waking up after a nightmare and had no memory what the nightmare was about. In such a case, one can be in the state of anxiety, but the state is not *about* anything. When this occurs the subject is anxious, but she is not anxious about anything in particular. The possibility of this case shows that it is possible for qualia to outstrip its representational content, and the fact that it actually occurs shows that qualia really do outstrip their representational content. Along the same lines, one can hold that many mental states do not have any representational content, although they do have phenomenal content. Even though anxiety, depression, and euphoria are all mental states with phenomenal content (there is something it is like to be in those states), it is difficult to accept that these states represent the world as being a certain way. These examples show that some phenomenal states have no representational content, which is to say that the phenomenal content outruns the representational content. Therefore, they show

[49] For a similar argument along these lines, see Rey (1998). It is worth noting that I have already committed myself to some aspects of conscious experience being non-conceptual (and thus not intentional) in §2.2.

that strong representationalism is false.

Strong representationalists will resist the counterexample presented by moods and emotions that appear to have phenomenal content without representational content by trying to show that these states do in fact have representational content.[50] They claim that although the mental state does not appear to be about anything, the state does in fact represent information about the subject. When a person experiences anxiety, depression, elation, and the like, there are changes in the person's physiology, which representationalists contend are being represented. Although I find this explanation inadequate, there is a problem that runs deeper than a clash of intuitions. The deeper problem with this approach is that it does not show the intentional content is part of the representational content of the conscious experience. So, even if we take these moods and emotions to represent states about the subject having them, this does not allay the concern that the phenomenal content is exhausted by the representational content.

My third argument against strong representationalism is based on the possibility of a mental state's having different phenomenal content while having the same representational content. According to strong representationalism, the phenomenal content of an experience is exhausted by its representational content. But if it is possible to have the same representational content with different phenomenal content, then there must be something more to

[50] For a standard way of making this argument see, Tye (1995), pp. 111-119, 125-131.

phenomenal content than its representational content—otherwise, the different phenomenal contents couldn't be different. Consider, for example, the visual phenomenal experience of a round ball and the tactile phenomenal experience of a round ball. With respect to the representation of a round ball, the representational content is the same in both states. However, there is a phenomenal difference between the visual and tactile experiences of a round ball. If strong representationalism is true, however, then it should not be possible to have the same representational content and different phenomenal content. But it is possible. So, strong representationalism is false.

Representationalists typically respond to this sort of objection by insisting that the representational content is different. The problem with this response is that while this may work in some cases, it will not suffice to cover every example of this sort. For example, the representationalist may plausibly say that the visual experience of a round ball represents both the shape and color of the ball, whereas the tactile experience represents the ball's shape and texture. But we can imagine a special sort of echolocation that can be used to represent the ball's shape and texture without the phenomenal content being identical to tactile experience. If this is so, then the representational content could remain the same while the phenomenal content differs. Therefore, there is something more to the phenomenal content of experience than its representational content.

Summing up, I believe that representationalism does not present a promising way to save physicalism from the KI. First, it is unmotivated since it

denies non-physical properties at the cost of embracing unreal objects or properties. Indeed, if one is concerned with overpopulating one's ontology, it seems hardly palatable to recognize a multitude of unreal objects for the sake of keeping one's metaphysical commitments uncomplicated. Second, even if one can motivate an account of physicalism that accepts unreal objects, the physicalist still needs a non-question-begging way to determine how one can tell which representational states are veridical and which ones are misrepresentations. Finally, I've presented two kinds of counterexamples to the strong sort of representationalism employed by Jackson to motivate the position that phenomenal color experiences systematically misrepresent the way the world is. Since there are good reasons to think that the phenomenal content of consciousness is not exhausted by its representational content, strong representationalism does not suffice to show that the KI is false. Therefore, representationalism in and of itself cannot be used as a plausible way for the physicalist to avoid the Knowledge Argument.

3.3 Concluding Remarks

The strong denial of the KI has been approached from both ends of the spectrum. First, there is Dennett's view that the KI is false by trying to trump up the extent of one's knowledge given that one knows all the physical truths. Second, there is Jackson's strong representationalism that downplays the knowledge that seems to fall outside the purview of truths deducible from

physical truths. Both approaches fail to circumvent the KI. However, this does not mean that the KI is safe. The strong denial of the KI may fail because, well, it is too strong. Since the falsity of physicalism follows from the inability to deduce propositional knowledge of conscious experience from the complete set of physical truths, it remains a possibility for the physicalist to admit that there are some kinds of knowledge (non-propositional kinds of knowledge) that are not deducible from the complete set of physical truths. This weaker denial of the KI can pay some homage to the intuition, while avoiding the Knowledge Argument. In the next chapter, I shall consider the most plausible ways for physicalists to reject the Knowledge Argument by mounting a weaker denial of the KI.

CHAPTER 4: A WEAK DENIAL OF THE KNOWLEDGE INTUITION, OR DENYING THAT MARY LEARNS A NEW PROPOSITION OR TRUTH

In the previous chapter I concluded that physicalists cannot plausibly avert the Knowledge Argument by claiming that the Knowledge Intuition (KI) is false in a strong sense. The KI is the intuition that when Mary leaves her black-and-white lab and experiences color for the first time, she comes to know a new truth or proposition. In other words, certain truths or propositions about the subjective character of conscious experience are not deductively implied by the complete set of physical truths. The strong denial of the KI states that Mary learns absolutely nothing new about the world when she experiences color for the first time.

A more plausible route for the physicalist to take is to accept that there is some sense in which Mary learns something new when she leaves the black-and-white lab, but that the sense in which Mary learns something new does not constitute her learning a new truth or proposition about the world. In this way, the weak denial of the KI can acknowledge that there is something new about Mary's epistemic state when she experiences color for the first time, but it still circumvents the Knowledge Argument because it undercuts the idea that Mary's new epistemic status constitutes propositional or factual knowledge about the world. In this chapter, I will argue against the two following attempts to deny the KI in this weaker sense: the ability hypothesis and knowledge by acquaintance.

4.1 The Ability Hypothesis

One of the earliest responses to Jackson's Knowledge Argument is the ability hypothesis. Proponents of the ability hypothesis claim that Mary does learn something new when she experiences color for the first time, but they maintain that her discovery is nothing more than her gaining new abilities.[1] According to this response, prior to her acquaintance with phenomenal redness Mary lacked the ability to remember the experience of phenomenal redness; Mary lacked the ability to imagine experiences with phenomenal redness; Mary lacked the ability to recognize other experiences of phenomenal redness as being of the same kind.[2] After her first acquaintance with phenomenal redness, Mary acquires new abilities: the ability to remember, recognize, and imagine with phenomenal redness. The contention of the ability hypothesis is that the gaining of these abilities is the *only* thing that Mary learns. The acquisition of a new ability, or know-how, however, is not propositional knowledge, or knowing-that. Since know-how is non-propositional, it does not count as propositional or factual knowledge about the actual world. So, the ability hypothesis maintains that Mary's discovery in no way threatens the truth of physicalism. In other words, Mary's acquaintance with phenomenal redness does not adequately

---

[1] This view is best known from Nemirow (1990) and Lewis (1988).

[2] Lewis advocates for all three abilities, whereas Nemirow takes only the ability to imagine as relevant. Although, what Nemirow includes among the ability to imagine might also include the ability to remember.

support the KI. David Lewis succinctly sums up the way the ability hypothesis is supposed to save physicalism:

> The ability hypothesis says that knowing what an experience is like just *is* the possession of these abilities to remember, imagine, and recognize. It isn't the possession of any kind of information, ordinary or peculiar. It isn't knowing that certain possibilities aren't actualized. It isn't knowing-that. It's knowing-how. Therefore it should be no surprise that lessons won't teach you what an experience is like. Lessons impart information; ability is something else. Knowledge-that does not automatically provide know-how.[3]

The ability hypothesis, if correct, explains why knowing what it's like cannot be learned through physical information. Since generally all the information of know-how cannot be communicated through objective propositions, it explains why knowing what it's like defies being communicated through the objective language of physics. Furthermore, the ability hypothesis naturally fits with the way we typically test whether someone does know what an experience is like.[4] For example, if you want to test whether people know what it's like to see the color mauve, your answer will likely hinge on their ability to imagine, remember, or recognize the color. Stated differently, if someone lacked the abilities to imagine, remember, and recognize the color mauve, you would be justified in concluding that the person doesn't know what it's like to experience mauve. These insights lead advocates of the ability

---

[3] Lewis (1988), p. 100.

[4] This point was brought to my attention by Alter (1998), p. 37.

hypothesis to believe that knowing what it's like to undergo certain phenomenal experiences is nothing more than the possession of certain abilities.

All parties in this dispute accept that Mary acquires new abilities by being acquainted phenomenal redness. So, the ability hypothesis does accurately describe an epistemic change in Mary after her first experience of phenomenal redness. The question is whether it satisfactorily accounts for *all* the epistemic changes implied by the thought experiments that undergird the Knowledge Argument. One problem for the ability hypothesis is that the abilities to remember, recognize, and imagine are neither necessary nor sufficient to account for Mary's new epistemic state. If either of these criticisms of the ability hypothesis is defensible, then the KI is not undercut by the ability hypothesis.

First, let's consider whether the abilities of recognition, remembrance, and imagination are necessary to account for Mary's new epistemic state. Following Earl Conee,[5] we can imagine someone who lacked the abilities of recognition, remembrance, and imagination (perhaps due to brain damage), and yet it still seems possible that this person can be acquainted with phenomenal redness. Despite lacking the aforementioned abilities, the disabled person can still know what it's like to experience phenomenal redness. As Conee explains, this example "enables us to see that knowing what an experience is like requires nothing more than noticing the experience as it is undergone. … In fact, no

[5] Conee (1994). Similar criticisms appear in Alter (1998), pp. 37-38 and Gertler (1999), pp. 322-326.

ability to do anything other than to notice an experience is required."[6] So, it is not necessary to have the abilities to recognize, remember, or imagine color experiences to account for Mary's new epistemic state.

Perhaps an advocate for the ability hypothesis may wonder, based on the final quotation from Conee in the previous paragraph, whether the ability to notice is a necessary condition for knowing what it's like to experience phenomenal redness. There are different ways to understand what is involved to perform the act of noticing, and the sense in which one takes the act of noticing will crucially reveal whether the ability to notice is necessary for knowing what it's like. The way Conee takes "noticing" is clearly an epistemic notion, one where to notice *p* is to believe that *p*. This sort of noticing does seem essential to knowing what it's like, but it involves propositional belief and therefore is no help to the ability hypothesis. There are, however, other senses of noticing, which do not involve propositional content. For example, the ability to notice could be understood simply as paying attention to something or the ability to focus on something. But these non-propositional senses of noticing seem to amount to nothing significantly different than a subject's acquaintance with a non-conceptual mental state.[7] So, Conee's sense of noticing involves propositional belief, which means it is of no use to defenders of the ability

---

[6] Conee (1994), p. 139.

[7] For more on this position, see the end of §4.2. There, I discuss Michael Tye's current position, which amounts to a weak denial of the KI by equating Mary's new knowledge with direct acquaintance with a non-conceptual state of mind.

hypothesis, and other senses of noticing that are supposed to be non-conceptual are not of use to the ability hypothesis.

Second, consider whether these abilities are sufficient to account for Mary's new epistemic state. Once again, Conee's work illustrates this point.[8] Imagine Martha who is capable of imagining new shades of color based on interpolations of colors that she has seen. Suppose Martha has never come to know what the shade of cherry red is like, but she does know with perfect recall what the shades of fire engine red and burgundy are like. Upon learning that cherry red is perfectly midway between fire engine red and burgundy, Martha is able to imagine what cherry red is like. Even though she is perfectly capable of knowing what this shade of red is like at the moment before she performs the imaginative interpolation, she remains ignorant as to what it's like to experience cherry red. Thus, the ability of knowing how to imagine a certain shade of color is not sufficient to know what it's like to experience that color.

In a recent essay, Laurence Nemirow has responded to these objections.[9] It will be worthwhile to assess whether his latest rejoinders provide any hope for the ability hypothesis to offer a way out of the Knowledge Argument for physicalism. Nemirow responds to the charge that the abilities of recognizing, remembering, and imagining are not necessary for knowing what it's like by arguing that we should feel a strong inclination to deny that any person who

---

[8] Conee (1994).

[9] Nemirow (2007).

lacks these abilities really does know what it's like to be acquainted with that color. In order to illustrate the sort of reasoning he has in mind here, I will quote Nemirow at length:

> When we attribute knowledge of what it's like to see tomato red [or a red tomato] to ordinary people who are staring at a red tomato, we assume that they can activate a panoply of imaginative abilities. For example, seeing a red tomato, I can compare its hue to other colors that I can visualize or remember; I can imagine that the redness of the tomato occupies a larger or smaller portion of my visual field than it does; and I can imagine variations on the tomato's redness. If I were unable to activate any such abilities, you should be reluctant to agree that I know what it's like to see tomato red [a red tomato]. So incapacitated, I would lack conscious awareness of the tomato's color. More generally, knowing what an ongoing experience is like by virtue of having that experience entails, if nothing else, conscious awareness of the experience, which itself involves the abilities to *reflect* upon the experience, not merely to endure it—or in the particular case of the tomato, to stare intently at its redness.[10]

Nemirow's response claims in part that lacking *all* the abilities would constitute grounds for denying that a person knows what it's like to be acquainted with a particular color. From the passage quoted above, we see that the only ability that he takes to be necessary in all cases of knowing what it's like is the ability to reflect on the experience. Presumably, then, he would not object to saying that someone who lacked the ability to remember (but who retained other abilities including the ability to reflect on the experience) would still be rightly judged to know what phenomenal redness is like.[11] And the same would

---

[10] Nemirow (2007), p. 35.

[11] If advocates of the ability hypothesis are not willing to grant this concession, their position becomes untenable. It is patently obvious that someone who lacks the ability to remember knows what redness is like while he is intently staring at a red tomato. For more on this point, see below.

go for any of the other abilities and combinations of abilities as long as the person retained the ability to reflect on the experience.

But the ability to reflect on the experience, I will argue, requires the subject to have propositional knowledge of the object of reflection. Notice in the passage quoted above that Nemirow identifies the abilities of reflection in the following way: "I can compare its hue to other colors that I can visualize or remember; I can imagine that the redness of the tomato occupies a larger or smaller portion of my visual field than it does; and I can imagine variations on the tomato's redness."[12] Many of these abilities involve propositional judgments, such as the ability to compare the properties of one color experience to another. When comparing one color to another, the subject necessarily believes a proposition such as that *burgandy is darker than cyan*. But even this propositional belief requires the subject to have beliefs with propositional content to constitute the beliefs about the individual character of each color (e.g., burgundy is like *this*; and cyan is like *that*). So, to the extent that Nemirow thinks the ability to reflect on conscious experiences in this way is necessary for the ability hypothesis, he will not avoid the consequence that Mary acquires new propositional knowledge.

A more abstract problem with Nemirow's latest proposal is that it is implausible and unmotivated. The critical place where this is evident is when he writes,

---

[12] Nemirow (2007), p. 35.

> If I were unable to activate any such abilities, you should be reluctant to agree that I know what it's like to see tomato red [a red tomato]. So incapacitated, I would lack conscious awareness of the tomato's color. More generally, knowing what an ongoing experience is like by virtue of having that experience entails, if nothing else, conscious awareness of the experience, which itself involves the abilities to *reflect* upon the experience.[13]

I feel no tug whatsoever to agree with Nemirow's claim that the ability to reflect on an experience is necessary for knowing what an ongoing experience is like. For example, one might know what an ongoing experience is like, but not be able to compare it to any other experiences if the person has had no other experiences to compare it to (or cannot recall any other experiences). The same goes for the reflective ability to imagine the experience differently. Even if someone were convinced that the phenomenal properties of his present experience could be no other possible way, this does not prevent him from knowing the character of the experience. It appears that the only reason to agree with Nemirow is if one is already committed to the ability hypothesis. But taken on its own, his proposal is unmotivated.

For good measure, I will also argue that the ability to know how to imagine a color is not sufficient to know what it's like to experience that color. Nemirow contends that once you provide enough abilities to Martha, then Martha's abilities are sufficient to confer knowledge of what it's like. In particular, Nemirow claims that in addition to the ability to visualize the color, Martha also needs the imaginative ability to remember the color in question. But

[13] Nemirow (2007), p. 35.

this has two manifestly false consequences. First, it implies that Martha can be staring at a red tomato, form the belief that the tomato appears phenomenally red (on the basis of her immediate conscious experience), and fail to know what it's like if she doesn't have the capacity to remember. Second, even if we equip Martha with the ability to remember on Nemirow's account, then, Martha wouldn't come to know what it's like until after she engages her abilities to remember the experience. Nemirow owns up to this implication when he writes, "Martha has the ability to know what it is like to see cherry red ***after*** she is able to remember visualizing cherry red."[14]

Perhaps more cautious advocates of the ability hypothesis would not require Martha to use her ability to remember the experience in order to declare that she knows what it's like to see cherry red. They might say it is good enough that she possesses the ability, not that she actually uses it. But then it appears as if the ability is doing nothing whatsoever in the account of knowing what it's like. So, having the abilities specified by Nemirow are not sufficient for knowing what it's like to undergo certain experiences. Perhaps the ability to remember the experience puts Martha in a position to know what it's like to experience cherry red, but this is not the same thing as currently having the knowledge of what it's like to have the experience.

In sum, the know-how of recognizing, remembering, and imagining with color concepts is neither necessary nor sufficient to account for the complete

[14] Nemirow (2007), p. 34. Emphasis added.

change in Mary's epistemic state when she is released from her black-and-white environment. Since knowing what it's like does not require the exercise of abilities, the ability hypothesis does not satisfactorily provide a way for physicalists to reject the Knowledge Argument.

But even if we grant that there is some necessary covariation between knowing what it's like and exercising certain abilities, there is still a problem with taking knowing what it's like to consist only in the exercise of certain abilities—namely, that the abilities do not adequately explain the subjective character of conscious experience. As Brie Gertler explains, "if knowing what it's like just is having the recognitional ability, then this knowledge can explain the ability in only a trivial sense, if at all."[15] To see that knowing what it's like is not explained (in a non-trivial way) by the know-how that accompanies the right abilities, compare the following two explanatory statements:

(A) I possess certain abilities to recognize, remember, and imagine the phenomenal character of color experiences *because* I know what it's like to experience the phenomenal character of color experiences.

(B) I know what it's like to experience the phenomenal character of color experiences *because* I possess certain abilities to recognize, remember, and imagine the phenomenal character of color experiences.

If the ability hypothesis is correct, then (B) should be seen as a non-trivial explanatory claim. However, (B) only appears to be true when the word

[15] Gertler (1999), p. 323. I originally struggled to find the right wording to make this objection against the ability hypothesis, and I am grateful to have discovered that Gertler stated this objection clearly and forcefully. See Gertler (1999), pp. 323-324, for her version to which I am indebted.

"because" is taken in an evidential sense, not an explanatory sense. Instead, proposition (A) appears to be true when the word "because" is taken in an explanatory sense. This means that (A) and (B) have different truth-conditions and different meanings. Thus, even if the possession of certain abilities perfectly correlated with knowing what it's like, we would still have good reasons to think that having certain abilities is not the same as knowing what it's like. The upshot, then, is that Mary's new knowledge of what it's like to experience phenomenal redness is properly understood to explain her abilities, not the other way around. So even if the possession of certain abilities were necessary for knowing what it's like, we would still have good reason for thinking that knowing what it's like is more than possessing the right know-how.

In sum, the ability hypothesis has two significant shortcomings. First, the possession of certain abilities is neither necessary nor sufficient for knowing what it's like, which implies that the two kinds of knowledge are not the same. Second, having the right abilities fails to explain one's knowledge of what it's like. In fact, a person's abilities to recognize, remember, and imagine are best explained by his knowing what it's like to have the specified conscious experience. For these reasons, the ability hypothesis is not a plausible defense for physicalists to resist the Knowledge Argument.

4.2 The Acquaintance Hypothesis

The second weak denial of the KI disallows that Mary's new epistemic state constitutes propositional knowledge of the world because the kind of knowledge Mary acquires is knowledge by acquaintance.[16] Like the ability hypothesis, advocates of the acquaintance hypothesis accept that Mary undergoes some sort of epistemic change when she becomes acquainted with phenomenal redness for the first time. According to the acquaintance hypothesis, Mary's epistemic change is consistent with physicalism because her new epistemic status does not involve any new propositional knowledge. What Mary lacks in the black-and-white environment is the personal experience of the properties that she knows immaculately under physical descriptions. Even though, they claim, Mary possesses all the propositional knowledge of what it's like to see red from her complete account of physical truths, she still isn't directly acquainted with the experience of phenomenal redness. This response requires knowledge by acquaintance to be different from know-how or propositional knowledge. Conee describes the approach this way:

> Acquaintance with an experience does not require having either information or abilities. Acquaintance constitutes a third category of knowledge, irreducible to factual knowledge or knowing how. Knowledge by acquaintance of an experience requires only a maximally direct cognitive relation to the experience.[17]

---

[16] The acquaintance hypothesis is probably best known from Conee (1994) and Churchland (2004). In the next chapter I will address a different kind of acquaintance hypothesis—one that relies on indexical knowledge to claim that Mary comes to know an old fact with a new proposition.

[17] Conee (1994), pp. 136.

As for Mary, the acquaintance hypothesis supposes that when she is acquainted with phenomenal redness, she becomes directly acquainted with a physical property that she knew (propositionally) through her education in physics while she was in the black-and-white environment. The only epistemic difference with Mary after she is acquainted with phenomenal redness is that she now knows first-hand what that property is like, but she doesn't learn anything propositional through her acquaintance with the property. Similarly, a person could learn all the propositional truths there are about Houston (without ever visiting the city—perhaps by reading a *very* thorough visitor's guide), and then become acquainted with the city by visiting it in person. In visiting Houston, he does not learn a new proposition about it. Rather, the person now becomes personally acquainted with an entity that he knew exhaustively before the visit. In the same way, Mary's discovery of phenomenal redness can be explained as her coming to know by acquaintance a property that she had previously known only by description.

Those who reject the acquaintance hypothesis will most likely not deny that one significant reason that Mary's epistemic status changes is due to her acquaintance with a new experience. In fact, my approach to defending the Knowledge Argument requires Mary to be acquainted with a new kind of property (see §2.2 and §2.3). So, my defense of the Knowledge Argument will not dispute the claim that Mary's new epistemic state is a result of her becoming

acquainted with the properties of conscious experience. The difficulty for the acquaintance hypothesis is to maintain that Mary does not come to know a new proposition as a result of her acquaintance with the properties of conscious experience.

The problem for the acquaintance hypothesis is brought sharply into focus by Martine Nida-Rümelin's Marianna thought experiment.[18] Recall that Marianna is a woman who has never experienced color, although her vision is normal and she believes so justifiably. Marianna agrees to participate in a psychological experiment where she is placed in a house where everything is colorfully and arbitrarily decorated, which is where she experiences color for the first time. Marianna, however, is not taught the names of these colors, nor is she allowed to see objects that she knows by testimony to be a certain color. At the end of the experiment, she is presented with four slides of clear examples of red, yellow, blue, and green, and she is asked which color sample she thinks corresponds with the color normal sighted people see when they look at the sky (under ordinary daytime conditions). We can imagine that Marianna answers by pointing to the red slide and proudly says, "I believe it is this one." Of course, she is wrong. It is not until Marianna leaves the experiment that she finally comes to know the truth that the sky appears to be blue in the phenomenal sense.

[18] Nida-Rümelin (1995). I have previously discussed this thought experiment in §2.1.

The point of the thought experiment is that Marianna (like Mary) discovers a new proposition about the way the world is.[19]

The Marianna case helps show the difficulty for the acquaintance hypothesis because it allows Marianna to be acquainted with the phenomenal properties, yet it illustrates that her mere acquaintance with phenomenal color concepts does not adequately account for her epistemic change. Both Mary and Marianna now can discover new propositions about the world because they now possess phenomenal concepts. In response to this concern, Conee suggests that Mary's acquaintance with phenomenal redness provides a different way for her to know or refer to the same propositions she could express using physical information in the black-and-white lab. Just as you can represent the same proposition using different languages or symbols, Conee maintains that the phenomenal experience is merely a new way for someone like Mary to represent the same propositions she knew before her release.[20]

The problem with this response is that if Conee takes the propositions expressed with phenomenal concepts to be identical to the propositions that are expressed using physical concepts, then it would be plausible to suppose that if Mary were exposed to phenomenal concepts in the same manner as Marianna, she would be able to tell that the phenomenal concepts express the same

---

[19] See Nida-Rümelin (1995) and (2008), §3.3.

[20] Conee (1994), pp. 145-147.

propositional content as their physical counterparts.[21] But it is implausible to think this is what would happen. Mary's reaction would be no different from Marianna's. Furthermore, to accept that Mary could identify the propositional content of her phenomenal beliefs as identical to her physical beliefs would deny that there is any epistemic difference between these belief states (once again, the model is expressing the same proposition in two different languages).[22] But this poses an unpalatable dilemma for the acquaintance hypothesis: either (i) phenomenal and physical concepts express the same propositional content (i.e., they mean the same thing), which implies that there is nothing new about Mary's epistemic state when she is acquainted with phenomenal redness for the first time; or (ii) phenomenal and physical concepts do not express the same propositional content (they don't mean the same thing), which implies that there is a new truth Mary comes to know when she is acquainted with phenomenal redness for the first time.

The first horn is untenable because it removes one of the central motivations for the acquaintance hypothesis—namely, that there is an epistemic difference in Mary's new state. However, if the propositional content is identical, then there is no epistemic difference since the two expressions of the same proposition would have the same content (just as expressing the same belief in

---

[21] A similar problem for Conee's position is raised by Alter (1998), pp. 39-40.

[22] Stated this way, the view begins to look like Dennett's response. See §3.1 for my criticisms of Dennett's response to the Knowledge Argument.

two different languages has the same content and meaning in both languages). Therefore, Mary should be able to figure out what it's like to experience phenomenal redness by reflecting on the content of the equivalent physical propositions, and there is nothing new for her to learn. The second horn is untenable because it entails that Mary learns a new truth or proposition, which the weak denial of KI rejects.[23]

There is another way to take the acquaintance hypothesis, which has been suggested by Michael Tye's most recent work.[24] Much like Conee's proposal, Tye emphasizes that Mary comes to know what it's like to experience the color red when she leaves her black-and-white environment; Mary knew *about* the phenomenal character of color experiences, she just didn't know it first-hand.[25] Tye distinguishes his version of the acquaintance hypothesis from Conee's by stressing that he takes knowledge by acquaintance to be non-conceptual and non-propositional. Tye describes his distinctive approach this way:

> What needs to be appreciated is that knowledge by acquaintance of an entity is a kind of non-conceptual, non-propositional thing knowledge. I know the shade red$_{29}$ simply by being directly acquainted with it via my consciousness of it. In the case of the phenomenal character of the experience of that shade, I know it in just the same way—by acquaintance. Our consciousness of things, both particular and general, enables us to

---

[23] In the next chapter, I will consider a version of the acquaintance hypothesis that concedes there is a difference in meaning but accepts that they refer to the same fact. This approach does not deny the KI, so it will be addressed separately.

[24] Tye (2009). This is obviously a departure from his previous approach to the Knowledge Argument.

[25] See Tye (2009), p. 132.

> come to have factual knowledge of them, but that consciousness is not itself a form of factual knowledge at all.[26]

Tye's acquaintance hypothesis appears to split the horns of the dilemma I raised for Conee's account. The dilemma stated that Mary's new knowledge by acquaintance either has the same content as the physical truths Mary new before her acquaintance with redness or not. Tye suggests that Mary's knowledge by acquaintance is non-propositional and non-conceptual, and hence lacks propositional content. In other words, knowledge by acquaintance, according to Tye, is completely unlike propositional knowledge. Thus, Tye averts the dilemma by rejecting that knowledge by acquaintance has conceptual and propositional content.

While Tye's maneuver avoids the aforementioned dilemma, it presents another problem for his account. Most significantly, it appears that Tye is confusing *knowledge by acquaintance* with the relation of *acquaintance*.[27] I have defended a traditional account of knowledge by acquaintance in §2.2, which takes knowledge by acquaintance to consist of three acquaintances: acquaintance with the fact that *p*, acquaintance with the belief that *p*, and acquaintance with the correspondence that holds between the fact that *p* and the belief that *p*. When Tye describes knowledge by acquaintance as non-conceptual and non-propositional, it appears to fit what the traditional acquaintance theorist

---

[26] Tye (2009), pp. 136-137.

[27] See Fumerton (2008), §1 for a number of ways this confusion has been made.

describes as the relation of acquaintance. But it is a mistake to confuse the relation of acquaintance with knowledge by acquaintance. As a result, Tye's proposed knowledge by acquaintance is fatefully exposed to the Sellarsian dilemma and the problem of the speckled hen (see §2.2).

Even if Tye's account of knowledge by acquaintance is almost certainly untenable, it still cannot avoid the implication that Mary comes to know a new truth when she is acquainted with phenomenal redness. When Mary is acquainted with phenomenal redness she is acquainted with a new property, and she learns (if nothing else) that the new property is exemplified. She also learns that the property of phenomenal redness is correlated with certain physical properties she knew in the black-and-white lab. For Tye to deny that Mary comes to know a property when she is acquainted with phenomenal redness for the first time is tantamount to claiming that Mary is not acquainted with anything when she is acquainted with phenomenal redness for the first time. On the other hand, if Tye maintains that Mary is acquainted with an old property when she is acquainted with phenomenal redness for the first time, then (i) his response fails to account for the sense in which Mary has some new knowledge, or (ii) Tye's response ultimately must resort to the old facts, new modes defense of physicalism.[28] So, the dilemma that faced Conee's version of the acquaintance hypothesis can be finessed to counter Tye's view.

---

[28] The second consequent is expressly what he claims not to be doing in his latest work. See, for example, Tye (2009), p. 132.

Finally, I would like to highlight a problem with all versions of the acquaintance hypothesis. This objection notes that the acquaintance hypothesis fails to explain why optimal knowledge of the phenomenal can only be attained directly, whereas optimal knowledge of typical physical truths does not require direct acquaintance.[29] This prompts the obvious question: what is ontologically different about phenomenal states that explains why in order to know these truths optimally the subject must be acquainted with them? The acquaintance hypothesis "emphasizes the epistemic disparity between uncontroversially physical features (including being a certain neurophysical state) and phenomenal features," criticizes Gertler, "but does nothing to explain this disparity."[30] The property dualist, however, explains that the epistemic difference is a result of an ontological difference. Unlike physical truths, phenomenal knowledge is about the properties that essentially characterize the ontologically irreducible and subjective character of consciousness. Contrary to Conee's claim that the acquaintance hypothesis is "metaphysically noncommittal,"[31] the epistemic dissimilarity cries out for a metaphysical difference to explain it. Without an explanation for why acquiring optimal knowledge between the physical and the phenomenal is essentially different, the acquaintance hypothesis actually highlights the main reason the Knowledge Argument leads to property dualism.

---

[29] A version of this objection is also found in Gertler (1999), pp. 326-328.

[30] Gertler (1999), p. 327.

[31] Conee (1994), p. 147.

So, the acquaintance hypothesis is not a viable option for physicalists to reject the KI. To the extent that Mary's acquaintance with phenomenal concepts seems to yield new knowledge, it implies that Mary comes to know new propositions. But as advocates of the acquaintance hypothesis downplay the epistemic novelty that accompanies Mary's acquaintance with phenomenal concepts, they fail to accommodate the intuition that Mary does undergo an epistemic change when she is released. Finally, the acquaintance hypothesis fails because it offers no explanation as to why phenomenal properties require acquaintance to be known optimally. But this underscores the very point of the Knowledge Argument—the epistemic difference is best explained by a metaphysical difference.

4.3 Concluding Remarks

Although physicalists may try to steer a middle path by acknowledging that Mary's first color experience leads her to acquire a new epistemic state without learning a new proposition, the compromise fails. Once the concession is made that Mary does acquire something new epistemically, we've seen that it is not plausible to explain Mary's new knowledge without recognizing that she learns a new truth or proposition. Since I've already argued that a strong denial of the KI is indefensible in chapter three, the physicalist is forced to accept that Mary's new knowledge consists in knowledge of a new truth or proposition. The final physicalist defense to be considered, then, is whether Mary's learning a new

truth is proper grounds to conclude that physicalism is false. I shall address this final response in the next chapter.

CHAPTER 5: AGAINST PHYSICALISM'S OLD FACT, NEW KNOWLEDGE DEFENSE

If my arguments in the previous chapters have been sound, I have shown that there is no plausible way for the physicalist to deny the Knowledge Intuition (KI)—the claim that knowing all the physical truths is not sufficient to deduce all the propositional truths about the world. The final strategy available for the physicalist to resist the Knowledge Argument is to grant the KI, but deny the Physical Knowledge Intuition (PKI), which claims that if knowing all the physical truths is not sufficient for knowing all the propositional truths about the world, then physicalism is false. Physicalists who appeal to this response recognize that when Mary leaves the black-and-white lab for the first time, it is inevitable that she will acquire new propositional knowledge. But, they claim, new propositional knowledge does not necessarily imply that there are new facts in the world, if, for example, the truth-makers for the new knowledge are good old physical facts.

The basic idea that underwrites this defense is that one fact can serve as a truth-maker for a number of different propositions. Beginning with a simple case, consider the truth-maker for the proposition that *Des Moines is the capital of Iowa*. What makes this claim true is the fact that Des Moines is the capital of Iowa. But this same fact also serves as a truth-maker for the proposition that *either Des Moines is the capital of Iowa or the moon is made of green cheese*. Although this proposition is different from the first (e.g., it has different propositional

content and truth-conditions), it is made true by the same fact. Of course, in this simple case, the second proposition is deductively entailed by the first. Setting that feature aside, physicalists who appeal to the old facts, new modes of knowing response to the knowledge argument believe that the same facts which make true the propositions known in Mary's black-and-white lab also make true the propositions not knowable in the black-and-white lab.

Common to these physicalist approaches to the Knowledge Argument is the acknowledgement that phenomenal concepts have a unique epistemic status compared to physical concepts. Physicalists who use phenomenal concepts in this way, then, attempt to provide a reason that phenomenal truths cannot be known by *a priori* deductions from the physical truths. While this may not afford the physicalist with a physical account of the properties of conscious experience, it does give the physicalist an explanation as to why there is a conceptual "gap" between our phenomenal and physical concepts. The reason that the existence of a gap between the phenomenal and the physical is supposed to be compatible with physicalism is because of the different conceptual schemes we must use to think about the phenomenal and the physical. Thus, the gap is conceptual or epistemic, not metaphysical, according to physicalists who take this approach. As a result, the physicalist is forced to balance the epistemic exceptionality of the phenomenal, while not admitting any distinct ontological status of the phenomenal. Daniel Stoljar has recently described the basic mechanics and merits of the phenomenal concepts defense of physicalism this way:

> If it [the phenomenal concepts strategy] is correct, perplexity about the relation between experience and the physical in the philosophy of mind derives largely from a conceptual mistake, rather than from a potentially chronic ignorance of the science, or from incoherence in the notion of experience, or from the inconsistency of that notion with known facts, or from a hard to articulate but powerful resistance to dualism. In short, the strategy offers a pleasingly deflationary account of what are perhaps the main problems in philosophy of mind.[1]

There are a variety of ways to attempt this defense, and I will group them under the following two classes: appeals to indexical propositions and appeals to recognitional concepts.

5.1 Appeals to Indexical Knowledge

The first approach I will consider for the phenomenal concepts strategy analyzes phenomenal concepts in terms of indexical belief that is not reducible to any other kinds of belief. According to these physicalists, it is the *sui generis* nature of indexical knowledge that is to blame for the apparent gap between the phenomenal and the physical. In order to appreciate this point, we must first rehearse some familiar distinctions about the language of belief reports.

5.1.1 Belief Reports *De Re*, *De Dicto*, and *De Se*

One of W. V. O. Quine's important contributions to the philosophy of language is his analysis of a certain ambiguity in belief reports.[2] One interesting feature of descriptions of beliefs (and other propositional attitudes) is that they

---

[1] Stoljar (2005), p. 471.

[2] Quine (1956).

create *opaque* contexts, which is to say that substituting equivalent terms within the scope of propositional attitudes does not necessarily preserve the truth-value of the claim. To illustrate the opacity of belief reports consider the following: Lois Lane believes that Superman can fly, but it doesn't follow that she believes that Clark Kent can fly—even though it is true that Clark Kent is Superman. Quine's concern involved an ambiguity in how to interpret belief reports. Here is my variation on Quine's example of an ambiguous belief report:

(1) Ralph believes that someone is planting a bomb.

The ambiguity is evident because the belief report can be understood in one of two different ways:

(2) Ralph believes that there are bomb-planters

or

(3) Someone is such that Ralph believes of him that he is planting a bomb.

The difference between (2) and (3) is striking. Almost everyone believes the thought attributed to Ralph in (2), but few people believe what is attributed to Ralph in (3). Indeed, the appropriate response to the belief given in (3) is quite different than the response to the belief in (2). Under normal circumstances, if Ralph sincerely believes (3), he is going to make efforts to report the person he believes to be planting a bomb to the proper authorities. On the other hand, Ralph's behavior does not necessarily change because of the belief attributed to him in (2). Quine also noticed the logical structure that differentiates (2) and (3). Formally, we can express (2) as

(2*) Ralph believes: $\exists x$(x is a bomb-planter)

and (3) as

(3*) $\exists x$(Ralph believes that x is planting a bomb).

Notice that the existential quanitifier in (3*) has a wider scope than the one in (2*). Importantly, the quantifier in (3*) is bound to a variable that occurs within the scope of "believes." This observation suggests a syntactical way to distinguish *de re* belief reports from *de dicto* belief reports: a belief report can be classified as a *de re* report if and only if it contains a free variable within the scope of an opacity verb that is bound outside the quantifier scope of the verb. The belief report is *de dicto* if and only if it is not *de re*. Statements like (3) are thereby categorized as *de re* belief reports, and statements like (2) are categorized as *de dicto* belief reports.

Another way to carve up the *de re*/*de dicto* distinction concerns whether it is semantically permissible to substitute *salva vertiate*—where substituting equivalent terms necessarily preserves truth-values. On the semantic conception, a belief report is *de re* if and only if it permits substitution *salva veritate*; otherwise it is *de dicto*. For example, if the person in (3) that Ralph believes to be planting a bomb is named Abdul, the truth-value is preserved by replacing "someone" in (3) with Abdul.

In addition to *de re* and *de dicto* belief reports, some philosophers also believe that there is a distinct third category of belief reports that pertain to

beliefs about oneself, or beliefs *de se*.[3] For example, suppose Rick is hospitalized with a bad a case of amnesia—so bad that he does not know his own name or his ailment, nor is he told so. As it turns out, Rick's case of amnesia is recorded in the local newspaper, which Rick reads with great interest. By reading the newspaper, Rick may believe *de dicto* that Rick has a case of amnesia. Furthermore, suppose that there is mirror across from Rick, which he mistakenly thinks is a patient across from him, and somehow Rick comes to believe that the man he is seeing in the mirror has amnesia. Then, Rick can believe *de re* of the man he is seeing that he has amnesia. Advocates of belief *de se* will emphasize that Rick still does not believe that *I have amnesia*, nor can he deduce this belief from the content of his aforementioned beliefs *de dicto* and *de re*. And when he finally learns that *I have amnesia*, he will have acquired some new kind of belief that is distinct from his prior beliefs *de dicto* and knowledge *de re*. For this reason, belief *de se* is taken by many to be an irreducible third category of belief report.[4]

5.1.2 Indexical Belief Content and the Knowledge Argument

Let's grant that indexical belief reports are a distinctive kind of belief report. How does appealing to indexical belief reports apply to the Knowledge Argument? Physicalists can accept that Mary learns a new proposition—one

---

[3] The landmark work articulating and defending belief *de se* is Perry (1979). Although many of the same ideas had been pioneered in earlier work such as Castañeda (1967).

[4] By parity of reasoning, similar arguments can be made for belief content that essentially includes the indexicals "here," "now," "this," and "that."

whose content involves an indexical belief report—which is made true by a physical truth-maker (something Mary knew *de dicto* before she is released), and thereby accept the KI without abandoning physicalism. In the example given above, Rick's belief *de se* has distinct propositional content from his belief *de dicto* and knowledge *de re*, but this distinct propositional content does not require us to introduce new truth-makers in order to make sense of Rick's new propositional knowledge when he learns *I have amnesia*.

In the same way, some physicalists contend that we can account for the KI by acknowledging that the conjunction of all the physical truths doesn't entail all the propositional truths about the world because knowledge of the physical truths by themselves does not include possessing any indexical knowledge. After all, it seems that when Mary experiences color for the first time, she learns something new that is inextricably linked to herself. Indeed, what she seems to discover is essentially subjective and thereby a plausible candidate for indexical knowledge.

Although there are numerous ways to account for the semantics of indexical knowledge and how precisely such an account can be used to aid the physicalist,[5] all of these approaches must make a common assumption. The common assumption is that Mary's new knowledge (and thereby any proposition about a phenomenal truth) essentially contains indexical belief

[5] For the most well-known attempts to use indexical belief content to block the Knowledge Argument, see McMullen (1985), Bigelow and Pargetter (1990), Ismael (1999), Perry (2001), and O'Dea (2002).

content. As David Chalmers plainly states, this assumption is false: "Mary's central new knowledge does not involve any indexical element."[6] To see this, we'll need to revisit the Marianna thought experiment.[7] Marianna is like Mary insofar as she has never experienced colors besides black, white, and shades of gray, although she is perfectly capable of doing so. Marianna is then placed in an environment that is brightly colored where she experiences color for the first time. However, she is not told the names of any colors, nor is she given any information she could use to figure out their names. Finally, Marianna is asked which color she thinks the sky appears to normal sighted people under normal conditions, and she is presented with four slides with paradigm examples of red, green, blue, and yellow. Marianna points to the red slide and declares she thinks it's that one. Of course she is wrong, and she won't come to know that the sky appears phenomenally blue until later.

Now according to physicalists who insist that propositions about conscious experience are a result of distinct indexical content, they must hold that Marianna's belief at the end of the thought experiment is expressed as Marianna's having an indexical belief (IB): [8]

(IB) For any red object O, if Marianna is visually presented with O, then she will be in a position to form an indexical belief that she could express while demonstratively referring to O (e.g., by pointing):

---

[6] Chalmers (2004), p. 285.

[7] I have explained the Marianna thought experiment in §2.1. This thought experiment is from Nida-Rümelin (1995).

[8] The following argument is derived from Nida-Rümelin (1995), p. 253.

"the sky appears to normal sighted people like *this*."

According to the physicalist who is appealing to indexical knowledge, Marianna's IB would be equivalent to the claim that she has a phenomenal belief (PB):

(PB) Marianna believes that the sky appears phenomenally red to normal sighted people.

However, IB and PB are not equivalent since it is possible for one to be true, while the other is false. If Marianna's color experience is perfectly inverted (and she is ignorant of her inversion), for example, then she will express PB by demonstrative reference to green objects (that appear red to her) and she will not be in a position to form the right indexical belief when presented with red objects. Another case where the equivalence between IB and PB fails can be seen if Marianna is blind, but she possesses an instrument that allows blind people to identify the color of objects on the basis of their surface properties. In this case, IB can be true, yet Marianna will fail to have the belief ascribed to her in PB.

Another reason to think that Mary's new knowledge is not merely a type of indexical knowledge is evident from the best candidates for Mary's new knowledge, which I've suggested earlier (in §2.1). Among the new things Mary discovers include: the property of phenomenal redness is exemplified; necessarily red is a color; necessarily red is darker than yellow; phenomenal redness is more like phenomenal greenness than like phenomenal sourness. These propositions are different from indexical beliefs. Especially in light of the

prior argument that PB is not equivalent to IB, the supposition that the content of all of Mary's new knowledge is a type of indexical knowledge does not seem remotely convincing.

For these reasons, then, it is not plausible to hold that the unique content of phenomenal belief is equivalent to the candidate indexical belief. Therefore, physicalism cannot rely on indexical knowledge to account for the way in which Mary learns a new proposition about a physical fact.

5.2 Appeals to Recognitional Concepts

A second way that physicalists have attempted to explicate how Mary could come to know a new proposition about an old fact is to identify phenomenal concepts as recognitional concepts. Brian Loar, one of the earliest and most influential proponents of the recognitional concepts approach,[9] describes the basic stratagem of the project when he writes:

> Phenomenal concepts are recognitional concepts that pick out certain internal properties; these are physical-functional properties of the brain. They are the concepts we deploy in our phenomenological reflections; and there is no good philosophical reason to deny that, odd though it may sound, the properties these conceptions *phenomenologically reveal* are physical-functional properties—but not of course under physical-functional descriptions. . . . it is quite coherent for a physicalist to take the phenomenology at face value: the property of *its being like this* to have a certain experience is nothing over and above a certain physical-functional property of the brain.[10]

---

[9] Loar (1997). Similar physicalist approaches that take phenomenal concepts to be recognitional in the same way as Loar also include Tye (2003), Carruthers (2004), and Levin (2007).

[10] Loar (1997), p. 227.

Loar continues, "Phenomenal concepts are conceptually independent of physical-functional descriptions, and yet pairs of such concepts may converge on, pick out, the same properties."[11] Since two conceptually independent descriptions may pick out the same property (or so the defender of the phenomenal concept strategy urges) it is possible that Mary's experience of phenomenal redness may provide the opportunity for her to acquire different propositional knowledge than she knew in the black-and-white lab; nonetheless both propositions pick out the same property or fact that she knew about in the black-and-white lab.

Before advancing my discussion of the phenomenal concepts strategy, I would like to highlight that physicalist approaches subsumed under this name operate under an assumption about the nature of meaning and reference that is accepted by many philosophers today.[12] I believe, however, that adopting this approach to meaning and reference is a grave mistake. In particular, I believe the error lies in the way in which the new theories of meaning and reference allow the meaning of a concept to be determined wholly by things outside of a person's mind. Consequently, I must engage in a protracted discussion of the philosophy of language. After illustrating how these theories are used to support the

---

[11] Loar (1997), p. 227.

[12] Alter (2006), §4 maintains that proponents of the old fact, new knowledge approach do not necessarily have to make this assumption. He explains, "one could argue that while the psychophysical conditional [i.e., the physical knowledge intuition] is *a priori* knowable by those who possess the relevant phenomenal concepts, Mary lacks those concepts before leaving the room." It is important to recognize, however, that every published defense of the old fact, new knowledge account that I know of does make this assumption.

phenomenal concepts strategy, I will provide at least a rough sketch what I think is fundamentally wrong in this approach to the nature of meaning and reference.

5.2.1 The New Theories of Meaning and Reference

The two most influential works that have shaped the new theories of meaning and reference are Hilary Putnam's "The Meaning of 'Meaning'"[13] and Saul Kripke's *The Nature of Necessity*.[14] These two accounts go hand-in-hand since they can be (and have been) combined to explicate how "meaning" is constituted by extra-mentality. We'll begin with Putnam's account of semantic externalism and then consider Kripke's causal theory of reference.

Putnam famously claimed that "'meanings' ain't in the head,"[15] ultimately because he argued that "meanings" are determined by extension. The extension of a concept is the class of objects which the concept picks out. For example, the extension of "water" for earthlings is the stuff that fills the lakes and rivers which is composed of H_2O. The intension of a concept, by contrast, is the internal content of a belief that a person has when he is referring to something. The intension of "water" does not necessarily include that water is H_2O. After all, people prior to 1750 (before the science of chemistry) had a concept of "water." Putnam uses the term "stereotype" to characterize the standard or normal

---

[13] Putnam (1975b).

[14] Kripke (1980).

[15] Putnam (1975b), p. 144.

descriptions of a natural kind.[16] So, the stereotypical features of water include that it is wet at room temperature, clear, odorless, etc. However, if the meaning of a concept is determined by its extension, then it follows that the meaning of "water" consists of its referent (and not its intension or stereotype). Since "water" does not refer to something in one's mind, it follows that the meaning of "water" exists outside the psychological purview of the subject on Putnam's view—hence the name semantic externalism.

The main motivation for Putnam's position is derived from his twin earth thought experiment.[17] Suppose in a universe far, far away there is a planet that is very similar to earth with inhabitants who are very similar to human beings. In fact, the only real difference is that the extension of the stereotype "water" on twin earth is not H_2O. Instead, on twin earth "water" is composed of stuff with some other chemical compound (let's abbreviate it as XYZ), which has all the macroproperties of the same stuff that we call "water" on earth. Now consider the differences between someone on twin earth who is thinking about "water" and someone on earth who is thinking about "water" (if necessary, assume that they both exist prior to the discovery of chemistry in their respective worlds). The brain states of the twin earthling and the earthling can be molecule for molecule identical; they could be "mental twins," if you like. But the extensions of their stereotypes are different. The twin earthling's stereotype picks out XYZ,

[16] Putnam (1975b), p. 147.

[17] Putnam (1975b), pp. 139-144.

and the earthling's stereotype picks out H_2O. The crucial question Putnam puts to his reader is whether "water" means the same thing for both the twin earthling and the earthling. Putnam believed it is obvious that they don't share the same meaning. On twin earth "water" means XYZ, and on earth "water" means H_2O. Since the intension of "water" appears to be identical, it must be the extension that accounts for the difference in meaning. Thus, "meaning" is extension or reference.

Putnam gives another example where virtually identical intensional states have different extensional classes, which he takes as further support for semantic externalism.[18] Suppose Hilary (a normal human on earth) does not know much about trees, and that he cannot tell the difference between an elm and beech tree. In other words, his stereotype of "beech tree" and "elm tree" appears to be the same in his idiolect. But clearly, "beech tree" and "elm tree" have different meanings—even Hilary knows *that*. So, if the stereotypes appear to be the same, and yet the meanings are different, the difference must be due to extension. This is taken to be another illustration of the maxim "'meanings' ain't in the head."

The final twist in the development of semantic externalism realizes that if meaning is determined by extension, then even the narrow intensional states that *appear* to be identical are in fact not the same.[19] If meaning is determined by

---

[18] Putnam (1975b), pp. 143-144.

[19] This move was first made by Burge (1982). Putnam's agreement is noted in Putnam (1996).

extension, then the apparently identical intensional states shared by the twin earthling and the earthling when they are thinking about "water" are not the same—they have different meanings and are thereby different states. Similarly, Hilary's stereotypes of a beech tree and an elm tree may appear to be indistinguishable by their intensional content, but the intensional content must be different according to semantic externalism because the referents are different. The upshot is that the intensional states are not the same, despite the fact that from the first-person perspective they are indistinguishable. Consequently, semantic externalism implies that optimal introspective reflection alone is not sufficient to understand the meaning of even the intensional content of one's own mental states.

Kripke has presented the position known as the causal theory of reference, where an internal state (such as a thought) is able to refer by virtue of its standing in the right causal relation to its referent. Of course, it is not a requirement of Kripke's account that one must believe that the constituents of one's own thought have been caused in the right way; it is only necessary that they do in fact stand in the right causal relation to their referent. "[I]t is not how the speaker thinks he got the reference," writes Kripke, "but the actual chain of communication, which is relevant."[20] Of course, Kripke wisely declines to provide the necessary and sufficient conditions that would be involved in

[20] Kripke (1980), p. 93.

constituting the right causal connection between thoughts and their referents.[21]

The next important aspect of Kripke's theory is that there can be identity claims that are necessarily true, yet not knowable *a priori*. On Kripke's theory the meaning of a name is not identical with any descriptive content; instead, names take their meaning according to the things they designate.[22] Furthermore, Kripke defined a designator as being a "rigid designator" if it picks out the same object in every possible world where that object exists.[23] One result for Kripke's account of reference is that names are rigid designators. So, the names "Hesperus" and "Phosphorus" pick out the planet Venus in every possible world where it exists. Consequently, when one believes that *Hesperus is Phosphorus*, he believes something which is necessarily true. And when someone believes that *Hesperus is not Phosphorus* that person believes something that is necessarily false. Consequently, modal truths about identity statements involving names are not knowable by *a priori* reflection on the names in question. This shows that Kripke's theory of reference allows for certain necessarily true identity statements to be *a posteriori* truths.[24] It is possible, in other words, for two names that rigidly designate the same object, to state a necessarily true identity, but for

---

[21] See Kripke (1980), p. 94.

[22] Kripke (1980), pp. 24-27.

[23] Kripke (1980), p. 48.

[24] Essentially, Kripke takes necessity and possibility to be metaphysical categories, and *a priori*city and *a posteriori*city to be epistemic categories. Thus, what falls within the purview of epistemic possibility (roughly, what is knowable by *a priori* reflection on one's thoughts) and metaphysical possibility (roughly, modal truths about the objects of the world) may diverge widely.

that necessary truth to be knowable only through empirical discovery. Kripke calls identity statements that are *a posteriori* necessary "theoretical identifications."[25]

Of course, there are many more interesting details about semantic externalism and the causal theory of reference that I do not have room to discuss. But I think I have provided the broad strokes that are crucial for how these views in the philosophy of language pertain to the phenomenal concepts strategy. I have highlighted the semantic externalist's tenet that "meaning" is extension or reference, and the causal theory's principle that names take their meaning from the objects that they denote. If meaning and reference are externalized in the ways outlined by these positions, then it is possible for seemingly unlike thoughts necessarily to mean the same thing. Take the names "Mark Twain" and "Samuel Clemens," for example. Although they both rigidly designate the same American author, it is possible for someone not to realize that "Mark Twain is Samuel Clemens." Likewise, physicalists who draw on the phenomenal concepts strategy are unmoved by Mary's inability to derive the meaning of "the tomato appears phenomenally red" from her complete knowledge of the physical truths. After all, there is no reason to suppose that theoretical identifications should be transparent to those who use the linguistic expressions meaningfully.

Although many physicalists have appealed to Kripke's theory of reference to explain how physical-functional concepts could rigidly designate the same

[25] Kripke (1980), pp. 140-144.

objects as phenomenal concepts (and thereby possess the same meaning), it is interesting to note that Kripke believed his theory actually provided a sound argument for dualism. Since responding to Kripke's argument plays an important role in developing the phenomenal concepts strategy, I will briefly rehearse his basic argument here. Let "*A*" name the phenomenal state of being in pain, and "*B*" name the physical-functional brain state that physicalists identify as being in pain. If *A* is identical to *B*, then *A* and *B* would pick out the same property in all possible worlds where the property exists. But, Kripke suggests, *A* and *B* do not pick out the same property in all possible worlds where *A* exists. The essential character of *A* is the phenomenal painfulness, and it is possible for state *B* to exist without there being any phenomenal character. As Kripke explains, "If *A* = *B*, then the identity of *A* with *B* is necessary, and any essential property of one must be an essential property of the other."[26] The key to other theoretical identifications being knowable *a posteriori* is that the names in such identifications do not directly pick out the essential properties of their referents. Since "A" picks out the property of being in pain by its essential property, any other attempt to pick out the property of being pain will do so through a contingent feature of pain, and thereby not yield a theoretical identity.[27]

---

[26] Kripke (1980), p. 148.

[27] Kripke (1980), pp. 152-153: "Pain, on the other hand, is not picked out by one of its accidental properties; rather it is picked out by the property of being pain itself, by its immediate phenomenological quality. Thus pain, . . . , is not only rigidly designated by 'pain' but the

The name given to phenomenal pain is unlike other names because it refers to its referent by directly picking out its essence. For future reference, let's introduce a new term, "*directly rigidly designate*," to characterize a name that refers by picking out its referent by its essential feature. By contrast, I will use the term, "*indirectly rigidly designate*," to characterize a name that picks out an object in all possible worlds where the object exists, but does so without picking out its referent by its essential feature. On Kripke's theory, identity claims that are *a posteriori* necessary require for at least one name to *indirectly rigidly designate* their referents.

Next, I will consider how advocates of the phenomenal concepts strategy make use of semantic externalism and the causal theory of reference to save physicalism from the Knowledge Argument. Along the way, I will touch on how physicalists reject Kripke's argument for dualism. At the end, I will consider some problems for the new theories of meaning and reference, which I take to undermine the phenomenal concepts strategy.

5.2.2 Phenomenal Concepts as Recognitional Concepts

As Loar spells out his version of the phenomenal concept strategy, he

---

reference of the designator is determined by an essential property of the referent. Thus it is not possible to say that although pain is necessarily identical with a certain physical state, a certain phenomenon can be picked out in the same way we pick out pain without being correlated with that physical state. If any phenomenon is picked out in exactly the same way that we pick out pain, then that phenomenon *is* pain."

takes phenomenal concepts to be recognitional concepts.[28] Loar specifies that recognitional concepts are type-demonstratives that are used to pick out certain objects, events, and situations by way of perceptual discrimination. Recognitional concepts fit the linguistic structure of "*x* is one of *that* kind." Other characteristics that Loar ascribes to recognitional concepts are that they are recognitional at their core, they need not involve a reference to a past instance, they do not depend on consciously accessible compositional analysis, and they are perspectival.

Phenomenal concepts and physical concepts, according to Loar, are conceptually independent, which is why knowing all the physical truths does not put one in a position to deduce all the truths about the world. Furthermore, both phenomenal and physical concepts can refer without either one doing so through a contingent mode of presentation. Consequently, Loar attributes the misstep in Kripke's argument and the Knowledge Argument to occur in what he calls the "semantic premise," which states as

> (SP) 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.[29]

In the terminology I noted above, the semantic premise states that theoretical

---

[28] Loar (1997), pp. 225-227. Other prominent views of phenomenal concepts are that they are quotational concepts (e.g., Papineau 2002, 2007) or token demonstrative recognitional concepts (Tye 1995). I will proceed as if phenomenal concepts are type demonstrative recognitional concepts, but my criticisms will equally apply to all formulations of the phenomenal concepts strategy.

[29] Loar (1997), p. 224.

identifications require at least one of the properties to be *indirectly rigidly designated*. If the semantic premise is false, then it is possible to accept that Mary comes to know a new proposition when she leaves her black-and-white room, but not give up physicalism. Loar explains it this way:

> if a phenomenal concept can pick out a physical property directly or essentially, not via a contingent mode of presentation, and yet be *conceptually independent* of all physical-functional concepts, so that Mary's history is coherent, then Jackson's and Kripke's arguments are ineffectual. We could have two conceptually independent conceptions of a property, neither of which connote contingent modes of presentation, such that substituting one for the other in an opaquely interpreted epistemic context does not preserve truth. Even granting that our conception of phenomenal qualities is direct, physicalism would not entail that knowing the physical-functional facts implies knowing, on an opaque construal, the phenomenal facts; and so the failure of this implication would be quite compatible with physicalism.[30]

The basic idea is that Mary's beliefs about the physical-functional brain states and about the phenomenal experience of redness (while both being identical) constitute an opaque context. For example, simply because it is true that Socrates believed that *rain is composed of water*, it doesn't follow that Socrates was in position to know that his belief also means that *rain is composed of H_2O*. Even granting that "water" means H_2O, the conceptual independence of water and H_2O doesn't permit us to assume someone who holds a belief about water is in a position to know his belief is also about H_2O.

Similarly, Loar claims it is invalid to assume that because Mary believes that *Smith's brain is instantiating a specific physical-functional process*, it doesn't

[30] Loar (1997), pp. 224-225.

follow that Mary is in a position to know that her belief has the same meaning as *Smith is having the mental state of experiencing phenomenal redness*. According to physicalists like Loar, phenomenal redness necessarily picks out the same property as a specific physical-functional brain state, but the modes in which they pick out that property are conceptually independent. Both the phenomenal concept and the physical-functional concept can mean the same thing, and yet even the best *a priori* reasoners may fail to see that the concepts have the same meaning (in the semantic externalist's sense of "meaning"). Therefore, identifying phenomenal redness with a specified physical-function process is not transparent, nor should we expect it to be so. If this is true, then Mary's new propositional knowledge does not constitute a problem for physicalism.

There are at least two sorts of responses to the phenomenal concepts strategy that may appear to create problems for the physicalist, but which can be quickly remedied by appealing to semantic externalism. First, it is tempting to argue that because the phenomenal mode of presentation *directly rigidly designates* its referent and the physical-functional mode of presentation *indirectly designates* its referent that Loar's defense is vulnerable to a weakness. The (SP) is false, claims Loar, because both the phenomenal and the physical modes of expressing their respective beliefs could do so essentially. There is some initial merit in thinking the following questions expose the basic problems with these claims: if there are two modes that non-contingently capture the essence of some property, shouldn't we be able to deduce *a priori* that these concepts pick out the

same property? If not, wouldn't this imply that one mode is lacking some feature that is captured in the other mode, which would suggest one mode isn't capturing the essence of the property in a non-contingent way? In other words, an argument can be mounted against Loar's position along these lines:[31]

(4) If a person understands a conceptual mode of expression that captures the essence of the property to which it refers, then the person is in a position to know the essence of the property to which the mode of expression refers by *a priori* reflection on the mode of expression alone.

(5) The phenomenal and physical-functional concepts that (allegedly) refer to the same property are conceptually independent; it is not possible to derive one concept from the other by *a priori* reflection alone.

Therefore,

(6) The conceptually independent phenomenal and physical-functional concepts cannot both capture the essence of the same property.

Of course, if (6) is true, then it spoils Loar's defense of physicalism since it would contradict his claim that two conceptually independent concepts can both capture the essence of a property. Loar clearly accepts (5), so he must deny premise (4). Loar rejects (4) on the grounds that it equivocates on the phrase "captures the essence of:"

> On one use, it expresses a referential notion that comes to no more than 'directly rigidly designate'. On the other, it means something like ' be conceptually interderivable with some theoretical predicate that reveals the internal structure of' the designated property. But the first does not imply the second. What is correct in the observation about rigid designation has no tendency to imply that the two concepts must be a priori interderivable.[32]

---

[31] An argument in this vicinity is suggested by Chalmers (2004), pp. 290-293.

[32] Loar (1997) p. 229.

Given semantic externalism, it is no surprise that two propositions under conceptually independent modes of presentation could rigidly designate the same referent, without the subject being aware of the modal truths that hold for these propositions. The allure of the unsound argument against Loar's position is thinking that the subject is in a position to know the essential property of the phenomenal state because the referent is given directly. However, if one grants semantic externalism to the physicalist, the physicalist is permitted to say that our grasp of the meaning of beliefs expressed by phenomenal concepts is no more transparent than Socrates's beliefs about water. In other words, simply because it seems like the phenomenal mode of presenting beliefs about pain *directly rigidly designates* its meaning and it seems like the physical-functional mode of presenting beliefs about pain *indirectly rigidly designates* its meaning, it doesn't follow that we are in a position to know the directness of either mode of presentation. All that the physicalist needs is the claim that it is possible for both meanings to *directly rigidly designate* the essence of the property to which they refer. And since semantic externalism is the thesis that our introspective awareness does not give us access to the meanings of our beliefs, it is the perfect means for averting this sort of anti-physicalist argument.

A second apparent problem with the phenomenal concept strategy is that the intrinsic features of each conceptual mode of presentation suggests that the best way to understand the strategy is to take the conceptual modes of referring

to the same property to be second-order properties of the property to which they refer.[33] After all, in order to distinguish conceptually independent concepts from each other, it is natural to conclude that they have distinct properties. Even granting that both the physical-functional and phenomenal concepts pick out the same property essentially by conceptually independent modes, the properties that make up these conceptually independent modes of referring need to be accounted for. Thus, the phenomenal properties themselves that constitute the phenomenal modes of presentation show that there is at least one new type of property in Mary's ontology the existence of which she did not know before.

Laurence BonJour has stated this objection recently with powerful rhetorical force in the form of a monologue from Mary's point of view that is worth quoting at length.

> You philosophers are really amazing! The idea that I a*lready* know the facts I am interested in—indeed all facts of that general kind—is simply preposterous. . . . If you suggest to me that there aren't really novel properties, but rather novel concepts or ways of representing or whatever, then (while finding that suggestion itself pretty hard to swallow) I would still insist that *which* concept or way of representing is involved in each case is still something that my physical knowledge doesn't give me a clue about. Perhaps, as you say, there some clever or complicated way in which the things I want to tknow are related to the physical things I want to know are related to the physical things I do know—maybe there is even some metaphysically necessary connection between them (assuming that it is kosher got materialists to believe in such things!). Anything like that, however, just *adds* to the list of facts that my physical knowledge doesn't reveal to me.[34]

---

[33] Examples of this type of argument can be found in Lockwood (1989), McConnell (1994), Gertler (1999), Chalmers (2004), Stoljar (2005), Chalmers (2007), Nida-Rümelin (2007), White (2007), BonJour (2010), and White (2010).

[34] BonJour (2010), p. 14.

Once again, this sort of argument need not concern the physicalist who accepts semantic externalism. Although Mary can learn a new proposition about color experiences through the phenomenal mode of believing, it doesn't follow that the properties which constitute her phenomenal mode of believing are not physical properties—even the very physical properties she knew beforehand. The physicalist can consistently maintain that the properties which constitute Mary's phenomenal mode of believing are physical-functional properties, which Mary knew in the black-and-white lab. But given semantic externalism, it is possible for beliefs about physical-functional properties and phenomenal properties necessarily to have the exact same meaning. So nothing interesting follows by noting that Mary couldn't derive *a priori* that phenomenal states are essentially the same as brain states. Once again, Mary's situation is parallel to Socrates's with regard to his beliefs about water. Socrates can't derive *a priori* that his thoughts about water mean the same thing as thoughts about H_2O, and Mary can't derive *a priori* that her thoughts about phenomenal states mean the same thing as thoughts about particular brain states. While this may seem extraordinary from the first-person point of view, to assume otherwise is to suggest that Mary has a grip on the essence or meaning of a concept because of her introspective access to her own psychology. But semantic externalism denies precisely this move, and thereby averts another attempt to derail the phenomenal concepts strategy.

In sum, the phenomenal concepts strategy that takes phenomenal concepts to be recognitional concepts can give the physicalist everything that is needed to stave off the Knowledge Argument, if the new theories of meaning and reference are true. Thus, in order to vindicate the Knowledge Argument, I must show that the new theories of meaning and reference are incorrect. Next, I will provide my basic reasons for thinking semantic externalism and the causal theory of reference are false. To strengthen my case, I will also sketch a defense of an alternative account of meaning and reference, which is of no use to the physicalist who intends to avoid the ramifications of the Knowledge Argument.

5.2.3 New Theories of Meaning Versus the Theory of Descriptions

The main problem with semantic externalism is that it grounds meaning in entities that are beyond the purview of the subject's grasp, and thereby it invites epistemological and regress problems. Previously, I have argued that the structure of justification must be foundationalist and employ direct acquaintance at its most basic level in order to prevent a number of regress problems (see §2.2), and the arguments I am raising in this section will parallel that reasoning in many ways. The best alternative to the new theories of meaning and reference, which I believe is the right theory of meaning and reference, is Russell's theory of descriptions.[35] Below, I will give a brief characterization of the theory of descriptions and explain why I think it is superior to semantic externalism and

[35] See Russell (1905), (1910), and (1912), ch. 5.

the causal theory of reference.

According to the theory of descriptions, thought is meaningful by virtue of its being constituted by things with which one is directly acquainted. Consider the following thought: (i) the most densely populated city on earth is in the northern hemisphere. According to the theory of descriptions, the thought expressed in (i) is meaningful because I am acquainted with its constituent parts, such as the properties of being dense, being a city, being populated, etc., and certain relations like being in and being the most. The acquaintance with these properties may be direct, or it may be reducible to a description that is constituted by other properties. The properties that make up those descriptions are, in turn, meaningful by virtue direct acquaintance or by being constituted by other descriptions that are ultimately grounded in properties and relations with which one is directly acquainted.

Another important part of Russell's theory of descriptions is that the meaning of names consists in being shorthand descriptions,[36] contrary to the causal theory of reference which takes names to be rigid designators whose meanings are devoid of descriptive content. For example, the meaning of "Julius Caesar" is not the person himself since that would be an extra-mental object with which we cannot be acquainted. Instead the meaning of "Julius Caesar" is whatever description comes before a person's mind when she thinks of Julius Caesar. The description may vary from person to person, but some candidate

[36] More precisely, the meaning of all singular referring terms are shorthand descriptions.

descriptions would include "the man who founded the Roman Empire," "the Roman leader who started a civil war by crossing the Rubicon," or simply "the historical figure who we call *Julius Caesar*" (in the last description *Julius Caesar* is the noise of the spoken words or the shape of the written words with which one is acquainted).

As I've emphasized above, an important aspect of Russell's views on description and acquaintance is that the very meaning of thought must be grounded in properties and relations with which we are directly acquainted. He declares unambiguously, "Every proposition which we can understand must be composed wholly of constituents with which we are acquainted."[37] Russell thought that if the contents of thought did not consist of things of which we are acquainted, it would result in a sort of skepticism that robs our thought of any significance. He wrote:

> for it is scarcely conceivable that we can make a judgement or entertain a supposition without knowing what it is that we are judging or supposing about. We must attach *some* meaning to the words we use, if we are to speak significantly and not utter mere noise; and the meaning we attach to our words must be something with which we are acquainted.[38]

Before stating the problems that plague the new theories of meaning and reference, I will quickly extol some of the familiar virtues of Russell's theory of descriptions.[39] One benefit of the theory of descriptions is that it can make sense

---

[37] Russell (1912), p. 58.

[38] Russell (1912), p. 58.

[39] For more on these and other virtues, see Ludlow (2009), §3.

of how propositional content that fails to denote anything can still be meaningful without introducing Meinongian non-existent objects and their apparent contradictory implications. To use Russell's famous example, we can express the meaning of the claims that involve negative existentials like, *Pegasus does not exist*, using a definite description without reifying a non-existent object. The theory of descriptions maintains that we can capture the meaning of the claim that *Pegasus does not exist* this way: it is not the case that (there is a unique *x* such that *x* is the winged horse fathered by Poseidon). The intentionalist alternative that was epitomized by Alexius Meinong's philosophy, required what the thought was about (in this case, Pegasus) to exist as the object of thought. The abominable result on this alternative is that there is something which is the object of thought (Pegasus), but it does not exist. Since the intentionalist alternative implies something contradictory (Pegasus both exists and does not exist), Russell's proposal is an austere way to make sense of our meaningful use of names that refer to nothing. Of course, intentionalists introduce different kinds of existence to dissolve the apparent contradiction in their theory. The virtue of Russell's account, however, is that we need not resort to extravagant metaphysical theories that postulate varieties of types of existence to account for negative existentials. After all, many people have the intuition that existence comes in one type—one where something unequivocally either exists or does not exist.

Without going into the details, it is also worth noting that Russell's theory

of descriptions also provides a way to make sense of fictional discourse without reifying the contents of fictional worlds, and it also presents a way to account for the semantics of *de re/de dicto* ambiguities of belief reports.[40] Of course, the descriptivist solutions to these problems can become complicated, but it is worth noting that any plausible solution to these problems will get complicated due to the subject matter they intend to explain. Given its ability to resolve a number of problems in the philosophy of language without complicating one's ontological commitments, it is no exaggeration to say that no theory of meaning and reference has accomplished so much with so little. Of course, this judgment stands pending that the objections to the theory of descriptions (which will be discussed later) can be adequately answered.

There are a number of problems with the new theories of meaning and reference, which I will present. First, I will argue that they fail to satisfy an epistemological regress of meaning. Then, I will raise a second set of problems, which aim to show that the new theories fail to account for various problems in reference fixing. Some of these include names that fail to denote anything, names that appear to have the right causal history and fail to refer properly, and reference involved in fictional discourse.

The first problem is that if meanings are not grounded in virtue of the subject's being directly acquainted with its fundamental constituents, then there

[40] Once again, see Ludlow (2009), §3.

is no ground for the subject to understand the content of his own thoughts.[41] The basic idea is that many of our thoughts succeed in picking out their referents and are meaningful because of our acquaintance with the properties and other predicate expressions that single them out. For example, it is possible for me to think about and refer to *my advisor's favorite movie*; or *the tallest female philosopher*. Although I'm ignorant about the identity of the tallest female philosopher, presumably I can meaningfully think about her and refer to her because of my acquaintance with properties like "being the tallest" and "being a female" and "being a philosopher." After all, if I wasn't acquainted with these properties, I couldn't use them to constitute the relevant thought. Likewise, even though I don't know the identity of my advisor's favorite movie, it seems that I can think about it and succeed in referring to it because of my acquaintance with the relevant properties and predicate expressions. So, I'm capable of thinking about and referring to certain things *indirectly*; that is to say, it is because I am acquainted with certain properties and predicate expressions I am capable of thinking about and referring to things like the tallest female philosopher and my advisor's favorite movie.

Suppose now that contrary to Russell's theory of descriptions, that there is no foundational thought which grounds meaning and reference by being known *directly*. If this were the case, it would be impossible to think about or refer to

---

[41] This argument has been developed by talking with Richard Fumerton and looking at a manuscript he is preparing for publication. After developing this argument, I discovered a similar argument has been proposed independently by White (2010), pp. 105-107.

anything. Without some foundational thought, we would be forced to say that all thought is meaningful by virtue of our grasp of other thoughts—all of which are also meaningful by virtue of our grasp of other thoughts. A little reflection on this supposition shows that it leads to a vicious infinite regress, which would prevent the possibility of anyone's having a thought that is meaningful. But since we succeed in having meaningful thoughts, it must be the case that there is no vicious regress and that there is some foundational thought that is known directly which underwrites meaning. It is precisely for this reason that Russell thought that if we failed to be acquainted with the fundamental constituents that compose our thoughts, our words would be insignificant or a mere noise.[42]

According to semantic externalism and the causal theory of reference, however, what stops the regress and gives thought its meaning is not something that is known directly. For these theories, what gives thought its meaning is something like its referent, cause, or whatever extra-mental thing that places meanings "out of the head." But if the thing that provides foundational meaning is outside the mind, then the regress does not terminate in something the subject understands or grasps. For example, if my thought about *the tallest female philosopher* is indirect or derivative, and what provides the foundational meaning from which it is derived are things outside the purview of my mind, then the extra-mental grounding does me no good to understand or grasp the meaning of my own thought. In other words, from my point of view, I would be in the exact

[42] Russell (1912), p. 58.

same position whether the extra-mental grounding existed or not. For this reason (and related reasons given in §2.2), it is necessary to ground thought in properties with which one is acquainted.

The semantic externalist may charge this line of reasoning as question begging. After all, the semantic externalist is not claiming that the regress goes on forever. Rather, semantic externalism grounds the foundation of meaning in a causal relation. Thought is meaningful, according to the semantic externalism, because it stands in the right sort of causal chain to its referent. Thought does have a foundation, claims the externalist, but the foundation does not necessarily terminate "in the head." My response is that grounding the meaningfulness of thought in this way is something which I find incredible and contrary to my own introspective reflection on thought. It seems clear to me that the content of my thought when it is meaningful is constituted by entities with which I am acquainted. The point of the example of indirect and direct thought is to show that we do, in fact, have an introspective stopping point for the meaningfulness of thought. In other words, the semantic externalist insists on a theory of meaning which seems completely contrary to our experience.

The second set of problems with the new theories of meaning and reference involve complications in fixing the reference of names due to either the theory's granting reference when it should not or the theory's inability to grant reference when it should. In a way, this criticism is related to the first. If meanings were grounded in foundational thought secured by direct

acquaintance, these problems would not arise. Thus, the problems I am raising here are a consequence of the externalization of meanings or tethering meaning to reference. The next set of problems, however, can be recognized without necessarily endorsing the Russellian theory of descriptions.

Consider, first, cases where a name has a causal history connected to one referent, but yet it comes to be associated with a different referent. Let's begin with a classic example, originally put forward by Gareth Evans.[43] When Marco Polo named the east African island "Madagascar," it appears that Polo misunderstood that the name was a native word that originally referred to a larger region of the continent of Africa that included Madagascar. Although Polo received the name "Madagascar" from a causal chain that is linked to the initial use of the name, he mistakenly applied the name to refer exclusively to the largest east African island. Since on the causal theory of reference it is only necessary for a name to satisfy the right sort of causal chain to its referent and Polo acquired the name "Madagascar" from a causal chain that originally referred to a larger continental region of Africa, it seems that the name "Madagascar" should refer to and thereby have its meaning as the land that constitutes the larger continental region of Africa, despite all appearances to the contrary.

A similar problem arises when one considers the causal ancestry for the name "Santa Claus." Although the name originally referred to a European saint

[43] Evans (1973).

(Saint Nicholas of Myra, a fourth century Greek bishop) and our current usage undoubtedly descends from this individual, it is wildly counterintuitive to suppose that "Santa Claus," as it is used today, refers to the European saint.

Of course, the defender of the causal theory will stress that it is not enough that there is any causal chain that connects a name with its referent; it must be an *appropriate* causal chain. Kripke tentatively suggests that what may be missing in cases like the names "Madagascar" and "Santa Claus" is that it must be part of the intention of the speaker to keep the same reference from whence he heard it.[44] This concession by itself is a serious blow to a pure causal theory of reference, and it invites those who wish to salvage some aspects of the theory to adopt a theory that is hybrid between the causal theory and the theory of descriptions.

Even worse, it isn't clear that these cases can be resolved in favor of the causal theory by showing that the speakers did not intend to use the words according to their predecessors. In Marco Polo's case, he intended to use "Madagascar" in accordance with the native reference, and there is no contradiction in thinking that the name "Santa Claus" descended in each of its causal links with the intention to satisfy the referent of the prior link. Of course, what is needed to prevent reference change of this sort—and what the causal theory of reference rejects—is that to secure reference across its ancestral "links" the meaning of a name must consist of its descriptive content, which remains

[44] Kripke (1980), pp. 96-97, 163.

essentially the same. In other words, it isn't enough to give the linguistic users in the causal chain of a name the right intentions to avoid reference change problems. Unfortunately for the causal theory of reference, some roads to reference change can be paved with good intentions.

A related problem arises when a name is given to what is believed to be a specific individual referent, but later it is discovered that the name has no unique referent. Science is chock full of examples of names of this sort. Consider, for example, the names "caloric," "ether," and "phlogiston," all of which were supposed to refer to entities, yet we came to discover that they do not refer to any unique referent. Most likely "ether" and "caloric" refer to nothing; and "phlogiston" turned out to be a combination of different elements. Similarly, one wonders whether some entities named in current scientific theories will turn out to refer to anything, such as superstrings, dark matter, and various names given to supposed subatomic particles. Some names of alleged persons also may have this status. Does "Robin Hood" pick out a person who existed in history or is he just a fictional legend? If "Robin Hood" existed, does the name "Robin Hood" refer exclusively to him or does it refer to a fictional legend whose meaning can significantly diverge from the alleged historical figure? And we could go on listing numerous examples of this sort. These names seem to create a problem for the causal theory of reference because it seems that these names are meaningful, yet many of them don't refer to anything at all, while for others it is not clear whether the origin of the name is tied to history or fiction (and thereby

whether it has a referent or not).

If names are supposed to be rigid designators and if the meanings of names are supposed to be their referents, how do the new theories of meaning and reference make sense of the fact that we can and do use names that fail to refer to anything in a meaningful way? Surely the scientists who tried to work out the theories that attempted to refer to "ether," "caloric," and "phlogiston," understood that these names have meaningful content. They weren't speaking nonsense when they talked about "ether," "caloric," and "phlogiston," and neither are we when we use these names today. When names fail to refer to anything, the new theories of meaning and reference are forced to say that these names are meaningless, or that they all mean the same thing (since they all refer to the same thing, namely nothing at all). From the Russellian perspective, these consequences stand as a *reductio* of their position. Since the theory of descriptions can account for the meaningfulness of these names, and how their meanings are distinct from one another, it is preferable to the new theories of meaning and reference.

Kripke himself has raised a puzzle about belief that illustrates a further problem with identifying a name's meaning with its referent.[45] Kripke gives the example of Pierre, a native Frenchman who speaks only French, who believes that "Londres est jolie," which expresses the same propositional content as the English sentence "London is pretty." Pierre, then, moves to London and learns

[45] Kripke (1979).

to speak English as a child in England would, rather than by translating French words to English (or using any way that relies on translating content from French to English). Somehow, Pierre never learns that "London" is the same places as "Londres." As the story goes, Pierre lives in a very ugly part of London, and he comes to believe that "London is not pretty." As a result, Pierre believes both that London is pretty and London is not pretty. Although Pierre believes contradictory things, he cannot be blamed for believing a contradiction without further information. The point Kripke raises is that this puzzle about belief is a genuine puzzle—one which he thinks any account of belief content must face.

The descriptivist, however, can account for the differences in Pierre's belief by unpacking the descriptive content that Pierre associates with the names "Londres" and "London." The fact that Pierre can simultaneously think both "Londres est jolie" and "London is not pretty" without being in contradiction is accounted by accepting that the meaning of a name is not fixed by its referent. The causal theory of reference, however, upholds that a name refers by standing in the right causal chain to its referent. The puzzle, then, that Kripke raises is a puzzle for those who hold that names refer in virtue of being caused by their referent in the right way. This problem vanishes, however, once we reject that notion that names are meaningful just by having the right causal relation to their referent.

Another untenable consequence of identifying meaning with reference

comes from Thomas Nagel, who has suggested that semantic externalism's failure to grant the meaningfulness of thoughts about radical skepticism (for example, the belief that "perhaps, I'm a brain-in-a-vat")—which has been celebrated as a virtue of semantic externalism[46]—demonstrates that semantic externalism is false.[47] If semantic externalism is true, then the belief "perhaps, I'm a brain-in-a-vat" is necessarily either false or meaningless. This follows because if I am a brain-in-a-vat who is being deceived about all of reality, then the reference for my terms (like "vat") doesn't pick out anything because the causal chain to its referent will be deviant (*ex hypothesi*) even if it is causally related to a unique object.[48] Putnam proudly commends his theory on this point when he writes, "although the people in that possible world [i.e., the brain-in-a-vat world] can think and 'say' any words we can think and say, they cannot (I claim) *refer* to what we can refer to. In particular, they cannot think or say that they are brains in a vat (*even by thinking 'we are brains in a vat'*)."[49] Nagel contends that theories like semantic externalism are "refuted by the evident

---

[46] See Putnam (1981). Undoubtedly, some proponents still take this to be an appealing feature of the theory.

[47] Nagel (1986), pp. 70-74. For a similar response see Fumerton (1995), pp. 45-47.

[48] The story is actually more complicated. For example, in order to make the example work in such a way as to sever any causal chain from thought to object, Putnam (1981), p. 6, suggests that "perhaps (though this is absurd) the universe just happens to consist of automatic machinery tending a vat full of brains and nervous systems." Recognizing these sorts of complications does nothing to assuage my convictions that these consequences stand as a *reductio* of the theory.

[49] Putnam (1981), p. 8.

possibility and intelligibility of skepticism."[50] Therefore, since the notion of global non-referring skeptical discourse is possible and (would be) meaningful, it follows that semantic externalism does not offer a correct theory of meaning.

Finally, let's consider how the new theories of meaning and reference treats the names of pure fictional characters, like "Sherlock Holmes," "Othello," and "Captain Kirk." Recall that on the causal theory of reference, names refer to the entity that they rigidly designate, and according to semantic externalism the meaning of a name is its referent. So, on the new theories, to whom do these names refer? Kripke himself is cagey about fictional discourse, but he seems to settle on the position that pure fictional names do not rigidly designate any person in the actual world, nor do they pick out any particular individual who exists in a possible-but-not-actual world.[51] Since there is no clear third possible way these names could refer to someone or something, it follows that these names have no referent and thus no meaning.

But this result is quite incredible. It is deeply implausible to suggest that we attach no meaning with the names of characters of pure fiction. If this were the only counterintuitive result for these externalist views of reference and meaning, perhaps I might be more forgiving. However, as the problems mount, we must be willing to look elsewhere for a theory that does not include these problems. And since the theory of descriptions can account for the meaning we

---

[50] Nagel (1986), p. 73.

[51] Kripke (1980), pp. 157-158.

attach to names of pure fiction, it is thereby preferable to the new theories of meaning and reference.[52]

Although the causal theory of reference and semantic externalism have their problems, the defender of those positions may insist that despite whatever problems they must grapple with, at least they aren't as devastating as the standard problems for the theory of descriptions. To round out my defense of the theory of descriptions, I will address several of the most cited problems and explain how the descriptivist can accommodate these challenges. I believe the final result shows the theory of description is still vastly superior to the new theories of meaning and reference.

One often cited problem with the theory of descriptions is that it presents a counterintuitive way to provide truth-conditions for propositions containing non-referring terms.[53] According to the theory of descriptions, propositions like *the present King of France is bald* turn out to be false, but many people find that ascribing a false truth-value to these sorts of claims doesn't harmonize with the way people ordinarily use language. By the evidence of ordinary linguistic usage, critics claim that the sentence should be deemed as neither true nor false, because perhaps it fails to present a complete proposition or it is defective in some other way. Russell's response noted that there are some cases where our

---

[52] It is important to stress that there are complications for *any* account of the meaningfulness of fictional discourse. The descriptivist faces these traditional problems, but he doesn't have to say that names of pure fiction are utterly meaningless, which is absurd.

[53] Originally due to Strawson (1950). More recently it has been raised by von Fintel (2004).

intuitions line up with the theory of descriptions (e.g., atheists who affirm that *God exists* is false), and that the support from ordinary language is hardly decisive.[54] Furthermore, defenders of the theory of descriptions point out that in ordinary language it is correct to say that it is false that *the present King of France is my doctor* or that *the present King of France is my housekeeper*.[55] So, this objection shows at best that stating some truths on the theory of descriptions turn out to violate rules for assertion according to Gricean conversational implicature,[56] but it is not sufficiently convincing to motivate reconsidering the theory of descriptions.

Another oft-cited problem for the theory of descriptions is that many descriptions associated with names are incomplete and thereby inadequate to pick out their referent or fail to constitute a definite description.[57] There are a number of ways this objection can be mounted. In ordinary circumstances, for example, someone might use the description "the cook" to try to refer to the person who prepared her meal. However, the description as it stands is ambiguous; it fails to pick out exactly one individual. Additionally this problem surfaces when the descriptive content for a number of names appears to be indistinguishable. For example, some people seem to associate the same

---

[54] Russell (1957).

[55] See, for example, Neale (1990).

[56] See Grice (1989).

[57] Initially raised by Strawson (1950) and Donnellan (1966). Some recent influential versions of this argument include Wettstein (1981), Reimer (1992), Szabó (2000), and Schiffer (2005).

description with the names "Tycho Brahe," "Nicholas Copernicus," and "Galileo Galilei," perhaps something like "the important sixteenth century European astronomer." Substitute, if you like, the descriptions associated with Egyptian pharaohs, like "Ramses," "Khufu," and "Djedefre" for a similar result among those whose descriptions for all three would amount to something like, "the ancient Egyptian ruler."

In response to the incompleteness of some descriptions, this can be remedied by acknowledging that some descriptions are elliptical, leaving the salient details to be determined by the context in which the description is given. Thus, thin descriptions such as "the cook" may be understood to have the expanded definition of "the cook who prepared such-and-such a meal at such-and-such a time and date," where the details are determined by the context or perhaps by indexicals.

For the second type of incomplete description, where the descriptions associated with numerous names appears to be the same, the descriptivist can plausibly maintain that despite the claims of their critics, there is unique descriptive content for those names. Among the descriptive content that most people possess for names like, "Ramses," "Khufu," and "Djedefre," is that *Ramses is not Khufu or Djedefre* and *Khufu is not Ramses or Djedefre* and *Djedefre is not Ramses or Khufu*. This also highlights one way the descriptivist can diagnose Putnam's example involving the meaning of "Beech Tree" and "Elm Tree." Putnam urged that in his idiolect it appears as if "Beech" and "Elm" mean the

same thing, and from that observation he concluded that the meaning must therefore come from something not "in the head." However, the descriptive content in Putnam's idiolect for "Beech" and "Elm" is different since he believes that *Beech trees are not Elm trees (and vice versa)*.[58] Some other ways for the descriptivist to account for these cases involves the strategy of incorporating the causal theory of reference into one's descriptions. This will be explained as a response to the next and final objection.

Advocates of the new theories of meaning and reference also have raised another family of objections for the theory of descriptions based on the apparent problems for the theory of descriptions to maintain stable and public reference as smoothly and successfully as the causal theory. There are at least two ways to think of this challenge. First, many people find the descriptivist's radical privatization of meaning to be utterly repulsive. For many who hold this position, it is extremely counterintuitive to take the meaning of names like "Napoleon," or "Shakespeare" to be held hostage to the subjective descriptions of each person. Instead, they take the meaning to be public and objective, which seems to favor the causal theory over descriptivism. A second way this challenge is important is that the continuity of singular referring terms across theory change in science, for example, seems to be preserved only on a causal theory of reference since the descriptions associated for entities, like "electrons," changes so radically that continuity seems to fail if meaning is determined by

[58] An argument along these lines is given in Searle (1983), pp. 200-205.

descriptive content.[59]

There are two complementary ways for the descriptivist to appease these sorts of intuitions. First, there is no reason why the descriptivist cannot incorporate the apparatus of the causal theory of reference into one's descriptions of terms that seem to have an objective or public meaning.[60] For example, the descriptivist is at liberty to suggest that part of the descriptive content for a name like "Napoleon" is that it is "the person who is called by that name who was the first link in a complex causal chain resulting in this use of the name." This is also a promising way to account for descriptions that appear to be ambiguous, such as "Khufu" and "Djedefre." Likewise, for specialized scientific terms, such as "electrons," one can plausibly take as an important part of one's description that electrons are "the entities that are the first link of a causal chain leading to the present scientific community's use of the name 'electron.' "

A second way for descriptivists to keep the meaning of names stable and to account for public meaning emphasizes that the descriptive content often consists of clusters of descriptions, and that sometimes these clusters have an order of prioritization. For example, someone might take the name "John McEnroe" to include among its descriptive content "the famous American tennis

---

[59] This challenge to scientific realism is due primarily to Feyerabend (1958) and Kuhn (1970). Some responses to the challenge explicitly invoke the causal theory of reference; see, for example, Psillos (1999), Boyd (2008), §4.1-4.2.

[60] This has been suggested by Fumerton (1988).

professional who won seven Grand Slam singles titles," "the tennis player who won nine Grand Slam doubles titles," "the men's professional tennis player who has a hot temper on the court," and "the most popular living person whose name is 'John McEnroe.' " Among this cluster of descriptions there are some descriptions that have a higher priority than the other descriptions. For example, if John McEnroe's actual name is not "John McEnroe" (it turns out to be a stage name), it would not essentially change the descriptivist's reference who prioritizes descriptions in such a way that the more essential descriptions succeed in describing and picking out its referent. After all, the reason we think John McEnroe's name is "John McEnroe" is because of his fame as a tennis player. Since some descriptions are parasitic on others, it is natural to take the less derivative description as more essential to fixing the descriptive content of a name than its derived contents. More essential to fixing our description of John McEnroe is that he is a famous tennis player, not that his legal name is "John McEnroe," or even that he won exactly seven singles Grand Slam titles.

I think this prioritazation of the clusters of descriptions can account for the stable and public use of scientific terms when it is combined with the descriptivist's incorporation of the causal theory of reference. What is central in all of the various descriptions given to, say atoms, is that they are the entities responsible for producing certain results, which led scientists to use the name 'atom.' " J. J. Thomson is famous for formulating the "plum pudding model" of the atom; Neils Bohr changed our understanding of the atom with something

like a "solar system model" of the atom; Schrödinger theorized that atoms behaved like waves; and James Chadwick revolutionized our understanding of atoms by discovering neutrons. These various descriptions may appear to be incommensurable and thereby threaten to break the continuity of reference for the term "atom." But if priority is given to the description most fundamental for fixing the meaning of the term "atom," then continuity of reference can be preserved. The description, "the entities responsible for the empirical results that led Thomson, Bohr, Schrödinger, and Chadwick to use the name 'atom,' " provides a way for the descriptivist to supply a stable description for the term "atom." Furthermore, since the other disparate descriptions of the atom are dependent on the empirical results of the relevant experiments, we can rightly prioritize this description over the conflicting descriptions given by the details of the different theories of the atom.

(Of course, I am not suggesting that all descriptions have a cluster of descriptions that can be prioritized in this way. All I am suggesting is that a defender of the theory of descriptions can claim that *some* terms can retain successful reference in the way described above.)

Admittedly, I haven't resolved all the objections to the theory of descriptions. However, I think I have presented a strong case for believing that it is most likely the correct view of meaning and reference. More importantly, I think that I have made the case that the causal theory of reference and semantic externalism cannot successfully account for the meaningfulness of thought and

reference.

The conclusion I have reached from this long excursus into the philosophy of language is that the physicalist cannot appeal to these new theories of meaning and reference to save physicalism from the Knowledge Argument. Yet without the causal theory of reference and semantic externalism, the phenomenal concepts strategy cannot plausibly respond to the Knowledge Argument, or even Kripke's argument for dualism. Since propositions about phenomenal properties derive their meaning from the subject's acquaintance with those properties, and the content externalist's attempt to make the meaning of these propositions to reside "outside the head," the physicalist cannot claim our introspective access to the content of thought—looking "in the head," so to speak, for the basis of meaning—is bound to mislead. Consequently, the admission that propositions about physical-functional states are different from propositions about the phenomenal content of conscious experience cannot be reconciled by the physicalist as having the same meaning or referent. So, as a corollary to the conclusion that the new theories of meaning and reference are false is the conclusion that the phenomenal concepts strategy cannot succeed in answering the Knowledge Argument.

5.3 Concluding Remarks

In this chapter, I've considered various ways for physicalists to accept the KI and deny the PKI. The first approach, which appealed to the indexical content

of belief failed primarily due to the fact that it is not plausible to account for all phenomenal belief in terms of indexical belief. The second version that appropriated phenomenal concepts failed because it required the adoption of the causal theory of reference and semantic externalism. Since the new theories of meaning and reference generate a vicious regress of meaning and have a number of other problems that are covered by the theory of descriptions, it follows that the new theories of meaning and reference fail in comparison to their most formidable rival. Therefore, the phenomenal concepts strategy fails as well.

It is worthwhile to take stock of the overall argument I've been assembling in this project. In chapters one and two, I've presented good reasons to accept both the PKI and the KI. In particular, in chapter two I presented both a *prima facie* case for the KI as well as a more substantial case based on a foundationalist epistemology grounded in direct acquaintance. In chapter three, I rebutted strong denials of the KI that deny Mary would learn anything new whatsoever when she left the black-and-white lab. Chapter four responded to weak denials of the KI, which tried to allow that Mary would learn something new when she left the black-and-white lab without learning a new proposition. And in the current chapter I've responded to physicalists who attempt to show it is consistent to affirm both physicalism and that Mary can learn a new proposition when she leaves the black-and-white lab. Given the plausibility I've marshaled for the PKI and KI and that the physicalist alternatives to these claims has been found insufficient to refute the Knowledge Argument, I take it that the

Knowledge Argument is a sound argument. Therefore, on the basis of the Knowledge Argument, we can conclude that physicalism is an unjustified position.

In what remains of this project, I am going to explore a topic closely related to the Knowledge Argument. Chapter six will explore some of the questions about the structure of the Knowledge Argument—particularly whether it is structurally self-defeating.

CHAPTER 6: DOES THE KNOWLEDGE ARGUMENT REFUTE DUALISM?

Some critics of the Knowledge Argument allege that even though the argument appears to succeed against physicalism, the same reasoning can be turned against any systematic metaphysical description of the world. An argument that refutes the most plausible rivals to one's own position is great, but an argument that refutes one's rivals as well as one's own position is not. The challenge to be addressed in this chapter is whether it is self-refuting for the dualist to affirm the Knowledge Argument.

Critics of this sort are taking an analogous approach to one way of understanding Gaunilo's criticism of Anselm's ontological argument for the existence of God.[1] On this reading of Gaunilo, he raises a counterexample to Anselm's argument (via the perfect island), but he does not point to a specific premise in Anselm's argument that is false. Rather, Gaunilo takes the counterexample to illustrate that Anselm's argument is wrong, even if he cannot show exactly where it has gone wrong. Similarly, critics of the Knowledge Argument who object to it on the grounds that it refutes dualism in addition to physicalism can be understood as claiming to cite a problem with the argument without claiming to know exactly where the argument has gone awry.[2]

In this chapter I will present the objection to the Knowledge Argument

---

[1] Anselm and Gaunilo (1077-1078).

[2] As we shall see below, some critic who raise this problem for the Knowledge Argument typically do have some idea of which premise they believe is the faulty one.

that contends the Knowledge Argument is problematic for both physicalism and dualism, and thereby unacceptable. Then, I will diagnose where the objection misfires and how dualists can resist being refuted by a similar type of Knowledge Argument.

6.1 The Charge of Self-Refutation

One way to understand this challenge to the Knowledge Argument is that it accuses the dualist of being saddled with an argument such that if it is successful against physicalism, then by parity of reasoning it also demonstrates that dualism is false. The dualist cannot cite the Knowledge Argument as a reason to accept his position, claims the critic, without exposing dualism to the same logic that is alleged to refute physicalism.

But how exactly does the Knowledge Argument apply to dualism? Below are two influential statements of the charge of self-refutation; the first quotation is from Paul Churchland and the second is from David Lewis.

> [A] long discursive lecture on the objective, statable, law-governed properties of ectoplasm [a short-hand for any non-physical substances], whatever they might be, would be exactly as useful, or *useless*, in helping Mary to *know-by-acquaintance* 'what it is like to see red', as would a long discursive lecture on the objective, statable, law-governed properties of the physical matter of the brain. Even if substance dualism were true, therefore, and ectoplasm were its heroic principal, an exactly parallel "knowledge argument" would "show" that there are some aspects of consciousness that must forever escape the *ectoplasmic* story. Given Jackson's antiphysicalist intentions, it is at least an irony that the same form of argument should incidentally serve to blow dualism out of the

water.[3]

> Let *parapsychology* be the science of all nonphysical things, properties, causal processes, laws of nature, and so forth that may be required to explain the things we do. Let us suppose that we learn ever so much parapsychology. It will make no difference. Black-and-white Mary may study all the parapsychology as well as the psychophysics of color vision, but she still won't know what it's like. Lessons on the aura of Vegemite will do no more for us than lessons on its chemical composition. And so it goes. Our intuitive starting point wasn't just that *physics* lessons couldn't help the inexperienced to know what it's like. It was that *lessons* couldn't help. If there is such a thing as phenomenal information, it isn't just independent of physical information. It's independent of every sort of information that could be served up in lessons for the inexperienced. For it is supposed to eliminate possibilities that any amount of lessons leave open. Therefore phenomenal information is not just parapsychological information, if such there be. It's something very much stranger.[4]

Both Lewis and Churchland make the same complaint against the Knowledge Argument. Their point is something like this: we can imagine that Mary studies and understands a comprehensive textbook of the completed metaphysics of dualism in her black-and-white room, and yet when Mary leaves the black-and-white room she will nonetheless still come to know something new when she sees a red tomato for the first time.[5] If physicalism is deemed false according to the Knowledge Argument because Mary learns something new when she leaves the black-and-white room, then dualism is falsified by the same criteria. Since it

---

[3] Churchland (2004), p. 168. See also Churchland (1985a) and (1985b) for other places Churchland has raised the self-refutation objection to the Knowledge Argument.

[4] Lewis (1988), pp. 93-94.

[5] *Cf.* Nagasawa (2002), p. 209 provides a story of "Mark" who is a dualistically omniscient scientist.

would be unfair to conclude that dualism is false on these grounds,[6] then as Lewis and Churchland imply, it would be equally unfair to conclude that physicalism is false.

It will be helpful to compare the structure of the Knowledge Argument to its close copy that is intended to target dualism. The claims I have identified as the premises that constitute the Knowledge Argument against physicalism are:[7]

(P1) If complete possession of all knowledge of physical truths isn't sufficient to provide all propositional knowledge of the actual world, then physicalism is false.

(P2) Complete possession of all knowledge of physical truths isn't sufficient to provide all propositional knowledge of the actual world, namely knowledge of the subjective character of conscious experience.

Therefore,

(P3) Physicalism is false.

The argument that Churchland and Lewis suggest would count against dualism would go something like this:[8]

(D1) If complete possession of all knowledge of the metaphysical truths of dualism isn't sufficient to provide all propositional knowledge of the actual world, then dualism is false.

---

[6] See Churchland (2004), p. 168.

[7] An alternative way of expressing the Knowledge Argument and how the charge of self-refutation is supposed to follow couches the argument in terms of an implicit application of Leibniz's law of identity of indiscernibles (*cf*. Jackson 1986). See Endicott (1995) for a reconstruction of these issues in that way.

[8] Endicott (1995), pp. 26-27, suggests a "conjunctive argument," that is to say that the problem is that if we give Mary exhaustive knowledge of the physical and non-physical via black-and-white television (for example), she still will not know all the truths about the world. On my construal of the problem, the dualist will have an answer to Endicott's version by giving a response to the argument posed by (D1)-(D3).

(D2) Complete possession of all knowledge of the metaphysical truths isn't sufficient to provide all propositional knowledge of the actual world, namely knowledge of the subjective character of conscious experience.

Therefore,

(D3) Dualism is false.

The challenge for the dualist who endorses the Knowledge argument is to show how the argument from (P1) and (P2) is sound, while giving a principled explanation as to whether (D1), (D2), or both (D1) and (D2) are false. I will give such a response in the next section.

6.2 No Self-Refutation

The objection to the Knowledge Argument that is based on the charge of self-refutation can be answered by noting how exactly the argument must be changed to be applied to dualism. Once the changes have been made and understood, it will become evident how the dualist escapes the charge of self-refutation, while consistently applying the Knowledge Argument against physicalism.

Since the entirety of Frank Jackson's response to the self-refutation objection is relatively short (back when he still accepted the Knowledge Argument), I will quote the whole of it:

> My reply is that lectures about qualia over black-and-white television do not tell Mary all there is to know about qualia. They may tell her some things about qualia, for instance, that they do not appear in the physicalist's story, and that the quale we use 'yellow' for is nearly as

> different from the one we use 'blue' for as is white from black. But why should it be supposed that they tell her everything about qualia? On the other hand, it is plausible that lectures over black-and-white television might in principle tell Mary everything in the physicalist's story. You do not need color television to learn physics or functionalist psychology. To obtain a good argument against dualism (attribute dualism; ectoplasm is a bit of fun), the premise in the knowledge argument that Mary has the full story according to physicalism before her release, has to be replaced by a premise that she has the full story according to dualism. The former is plausible; the latter is not. Hence, there is no "parity of reasons" trouble for dualists who use the knowledge argument.[9]

Overall I think Jackson's response is correct, although I believe it is important to unpack and defend some of the crucial claims he makes in this short passage. It may be possible for a dualist to reject (D1) as the faulty premise, but Jackson's response seems to concede that the dualist need not reject (D1). I concur. So the focus of rebutting the charge of self-refutation will question the legitimacy of (D2).

It is probably worth noting that both Churchland and Lewis thought that once we disambiguate the notion of "knowing the physical truths" such that it includes know-how and know-by-acquaintance, then it becomes clear how the Knowledge Argument fails to apply to both physicalism and dualism.[10] Likewise, Ronald Endicott blames the Knowledge Argument's applicability to both physicalism and dualism on the indexical nature of her new belief.[11] Rather

---

[9] Jackson (1986), p. 55.

[10] See Lewis (1988) and Churchland (2004). Lewis believed that know-how alone was sufficient to account for Mary, whereas Churchland seems to be open to both know-how and knowledge by acquaintance to account for the Mary case.

[11] Endicott (1995), especially pp. 23-24.

than revisit the problems with these suggestions, I am going to presume that what I have said about these physicalist responses in chapter 4 and §5.1 has sufficiently shown that these physicalist responses are insufficient to reconcile Mary's new knowledge with the physical truths she learned in her black-and-white environment. For the purposes of the current chapter, then, I will not let physicalism off the hook for these reasons.

Additionally, it has been alleged that *some* forms of dualism are open to the charge of self-refutation.[12] For example, if one gives an account of dualism where phenomenal properties are constituted out of imperceptible atomic non-physical properties, it may be open to the charge that knowing all of the truths about the fundamental non-physical atomic bits fails to confer full knowledge of all the truths about the world. Interestingly, some forms of panprotopsychism, appear to fit this characterization of dualism.[13] Defined roughly, panpsychism is the philosophical position that *mind* (or conscious experience) is a fundamental category of reality, and that everything that exists possesses this aspect of reality to some degree. Panprotopsychism, in contrast, is the position that some type of (non-physical) property that constitutes minds or conscious experience is a fundamental category of reality, and that everything in reality possesses this kind of property to some degree.

---

[12] See Nagasawa (2002), pp. 210-212.

[13] Notable advocates of panpsychism include Whitehead (1929), Nagel (1979), Sprigge (1983), Griffin (1998), Rosenberg (2005), and Skrbina (2005). A notable advocate of panprotopsychism includes Chalmers (1996), pp. 276-310, and *may* accurately characterize Russell (1921), Russell (1927), Maxwell (1978), and Stoljar (2001).

Perhaps of most interest is the type of panprotopsychism that is suggested in the work of David Chalmers.[14] Chalmers neatly encapsulates his view when he writes, "wherever there is a causal interaction, there is information, and wherever there is information, there is experience."[15] He goes on to entertain a position that he cannot rule out where "simple systems do not have phenomenal properties, but have *protophenomenal properties*."[16] He describes protophenomenal properties as "properties more fundamental than phenomenal properties from which the latter are constituted."[17] This approach has the benefit of not conferring phenomenal experiences to simple systems, like thermostats, and it fits the description of panprotopsychism given above.

In the context of the Knowledge Argument, it seems that the panprotopsychist reading of Chalmers's account is subject to the charge of self-refutation. Since panprotopsychism is the view that phenomenal properties are constituted out of non-physical properties that satisfy the right causal-functional roles, it seems right to affirm that Mary could learn all about these non-physical protophenomenal properties and their causal-functional roles and still not know what it's like to experience phenomenal redness.

But the charge of self-refutation does not stick to Chalmers's suggestive panprotopsychism. The reason is that it is not possible to know the intrinsic

---

[14] Chalmers (1996), pp. 275-310.

[15] Chalmers (1996), p. 297.

[16] Chalmers (1996), p. 298.

[17] Chalmers (1996), p. 298.

nature of protophenomenal properties, if they exist. Chalmers explains that there is a skeptical price that comes with embracing panprotopsychism: "the cost is the postulation of a class of unfamiliar properties that we do not understand."[18] Thus, it is not possible to know the essential truths of this form of panprotopsychism, which is why it is not subject to refutation by the Knowledge Argument and physicalism is.

Even if panprotopsychism is subject to self-refutation, I have no interest in defending panpsychism or panprotopsychism. So, I am content to note this problem and move on. In short, the sort of property dualism that avoids these problems is one that does not have a reductive structure that is analogous to the reductive ontology offered by physicalism.[19] Property dualism that is not modeled on the reductive ontology in physics—in other words, a robustly non-reductive dualism—does not face the charge of self-refutation by appealing to the knowledge argument. This point has been made by Howard Robinson:

> 'Mental substance' is not something composed of 'ghostly atoms'—whatever that would mean—but something that is not made of anything at all. In so far as it has a structure, that structure would be entirely psychological—that is, would consist of the faculties, beliefs, desires, experiences, etc. There would be no autonomous sub-psychological stuff. Such a notion faces many problems, of course, but this is the Cartesian conception, not the ectoplasmic one; and against this conception the knowledge argument is irrelevant.[20]

Returning to Jackson's response, his central point is that the dualist is not

---

[18] Chalmers (1996), p. 298.

[19] See Nagasawa (2002), pp. 212-215.

[20] Robinson (1993b), p. 183.

in the same position as the physicalist because "lectures about qualia over black-and-white television do not tell Mary all there is to know about qualia"; whereas "lectures over black-and-white television might in principle tell Mary everything in the physicalist's story."[21] The problem with the analogy suggested by Churchland and Lewis, then, is that fundamental truths about dualism (such as the intrinsic character of phenomenal experiences) cannot be learned in Mary's black-and-white environment, while there is no reason to suppose that learning essential truths about physicalism requires the subject to experience color or to have visual sensations at all.

Recall that among the propositional truths that Mary comes to learn when she leaves the black-and-white environment include claims like that phenomenal redness is exemplified; that necessarily red is a color; that necessarily redness is more like yellowness than sourness.[22] In order to have a systematic understanding of these kinds of truths, however, it is not enough for Mary to learn about these properties within the limits of black-and-white experience. In other words, Mary can know the intrinsic nature of physical truths in a complete black-and-white environment, while it is not true that she can learn the intrinsic nature of non-physical truths in a complete black-and-white environment. In order to understand the content of the relevant phenomenal truths, Mary would need to be acquainted with the non-physical properties that constitute those

[21] Jackson (1986), p. 55.

[22] See the end of §2.1.

truths—and that is something she will not acquire from any amount of studies conveyed through black-and-white monitors.[23]

So, the dualist's response to the charge of self-refutation is that there is a significant difference between (P2) and (D2). (P2) is something that the Mary thought experiment supports, whereas (D2) is not upheld by the same considerations. The reasons for thinking that Mary cannot know all the non-physical truths (and thereby reject D2) cannot be used to support physicalists who believe that Mary cannot know all the physical truths (and thereby reject P2).

Since (P2) and (D2) are not supported by similar considerations, the physicalist could try to motivate the claim that the Knowledge Argument backfires against the dualist with its faulty support for (P1). Perhaps the best way to show that (P1) is unacceptable is to appeal to a brand of physicalism that does not uphold the physical sciences as the ground-level language to which consciousness must be reduced. In other words, the strategy for the physicalist is to defend an account of physicalism that incorporates the irreducible, subjective character of consciousness among the fundamental predicates in a completed physics. The target of the Knowledge Argument is, according to physicalists of this stripe, any attempt to provide a systematic, objective account of the mind. Perhaps the most influential physicalist who takes this non-standard approach is

[23] See my argument in §5.2 to the effect that acquaintance with the properties that constitute a proposition is necessary to know it.

John Searle.[24]

John Searle writes in many places that the ontology of consciousness is essentially subjective.[25] Searle contrasts the first-person ontology of consciousness with the third-person ontology of mountains and molecules. The ontology of consciousness is essentially subjective because a necessary condition for conscious experiences to exist is that a person must experience them. By contrast, mountains and molecules can exist without someone experiencing them. So, the ontology of consciousness is essentially subjective in a way that the ontology of mountains and molecules is not. In response to challenges like the Knowledge Argument, Searle claims that too much is made of the objective-subjective distinction. Physical truths are said to be purely objective, whereas truths about conscious experiences are taken to be subjective, and from this people conclude one is physical and the other isn't. Rather than admit that this concession places truths of consciousness outside the physicalist picture, Searle suggests that we should re-think what counts as a physical science.[26] He writes,

> the scientific requirement of epistemic objectivity does not preclude ontological subjectivity as a domain of investigation. There is no reason whatever why we cannot have an objective science of pain, even though pains only exist when they are felt by conscious agents. The ontological subjectivity of the feeling of pain does not preclude an epistemically

---

[24] See, however, R. J. Howell (2009), which advocates "subjective physicalism." This sort view may also be attributable to Russell (1921), Russell (1927), Maxwell (1978), and Stoljar (2001).

[25] For example, Searle (1992), pp. 93-100; Searle (2002), pp. 40-44; Searle (2004), pp. 94-95.

[26] A similar idea is Thomas Nagel's proposal to develop an "objective phenomenology." See Nagel (1974), pp. 449-450.

objective science of pain.[27]

Searle urges it is a blunder to conclude that consciousness has no place in an objective science because of its irreducibly subjective character.

Robert J. Howell reaches a similar conclusion from his assessment of the Knowledge Argument.[28] After presenting the Knowledge Argument and surveying physicalists responses to it, Howell's conclusion unequivocally states, "To the extent that physicalism claims physics is an objective theory and can completely describe the world, physicalism is shown false by the knowledge argument against objectivism."[29] Rather than conclude that physicalism is false, however, Howell draws a different lesson from this conclusion. The real target of the Knowledge Argument, claims Howell, is the attempt to provide a complete objective account of the world, which could apply to physicalism or dualism. According to Howell, the lesson to be taken from the Knowledge Argument is that any systematic account of the ontology of our world must include among its fundamental constituents the properties that are essentially subjective. "Only a theory that is in part subjective—in the sense that it draws essentially from the understanding one gets by actually having experiences," writes Howell, "can provide the complete story about the world."[30] In essence, Howell is proposing to re-conceive the nature of physicalism so that the

---

[27] Searle (2002), p. 43.

[28] Howell (2007).

[29] Howell (2007), p. 170.

[30] Howell (2007), p. 170.

irreducibly subjective character of consciousness is taken as fundamental.

The relevance of the position staked out by Searle and Howell is that it represents the most plausible way to understand why a physicalist would reject standard interpretations of (P1), because it fails to include the ontologically irreducible qualia among the catalog of things "physical." Physicalists like Searle and Howell are ready to concede that the Knowledge Argument refutes a kind of physicalism, namely one that excludes the properties of subjective experience as being irreducible and physical. But they challenge the assumption that an ontologically irreducible property of subjective experience cannot be a physical property. Hereafter, I'll follow Howell in referring to this position as "subjective physicalism." All of this naturally raises the question: is it consistent to be a physicalist and admit that there are ontologically irreducible properties of subjective experience?

Perhaps the central problem with subjective physicalism is that it betrays so many commitments of physicalism and concedes so many points non-physicalists cherish that it amounts to a position that is physicalist in name only. Physicalism has been motivated by a number of principles. One motivation for physicalism comes from the supposed benefits of simplicity or making a minimal number of ontological posits.[31] The basic idea is that there is really no need to expand one's ontology beyond the physical. The ontology revealed in the

[31] See, for example, Quine (1948), Smart (1959), Churchland (1988), and Poland (1994), p. 26.

physical sciences are taken to be the metaphysical foundations on which the rest of the world has been constructed. Appeals to non-physical substances and properties are characterized as unnecessarily complicating and bloating the ontological catalog of the world by those who accept this first motivation.

Another motivation for physicalism is the idea that there is an epistemic primacy for the deliverances of the physical sciences over all other sources of knowledge, including introspective or phenomenal knowledge.[32] Many physicalists have urged that there is epistemic security in the physical sciences, which must be the starting point for all other knowledge about the world. For this reason Daniel Dennett includes among his "ground rules" the rule of "No Wonder Tissue Allowed," which he describes in part as the attempt "to explain every puzzling feature of human consciousness within the framework of contemporary physical science; at no point will I make an appeal to inexplicable or unknown forces, substances, or organic powers."[33] Thus, the traditional methodology for physicalists is to begin with our knowledge of the physical sciences and then to know and explain other phenomena from that basis.

While there may be other considerations that motivate physicalism, these two are among the most prominent, and subjective physicalism does not fit with either of these traditional motivations. With regard to ontological simplicity or

---

[32] This perspective is exemplified by Churchland (1986), Dennett (1991), especially pp. 39-42, Wilson (1998), Kornblith (2002), and Stalnaker (2008).

[33] Dennett (1991), p. 40.

consilience, subjective physicalism quickly abandons a minimalist ontology of the physical sciences, and increases the number of ontological posits by adding the irreducibly subjective character of conscious experience to the catalog of properties in the world. Once physicalism gives up the first motivation and permits among the fundamental ontological categories properties that are not among the traditional physical sciences (such as subjective phenomenal properties), the alleged parsimony of the physicalist picture is significantly weakened, if not altogether relinquished. What is the ontological difference, after all, besides the title "physicalism", in property dualism and subjective physicalism?

Subjective physicalism also fails to meet the second motivation for physicalism. I argued in §2.3 that since the phenomenal is epistemically privileged, it follows that the subjective character of conscious experience cannot be deduced from physical knowledge. In fact, if the phenomenal is given epistemic priority over the physical, the upshot is that our knowledge of the physical world (if there is any such knowledge) must be indirectly known through the phenomenal. Subjective physicalism concedes that the phenomenal is both epistemically and ontologically privileged. This undermines the physicalist motivation that prioritizes scientific knowledge in an attempt to know other features of our world. Indeed, if it is followed to its logical terminus, the privileging of our knowledge of the phenomenal results in a view where the existence of the physical world must be justified and the existence of the

phenomenal is most secure. So, rather than trying to fit the phenomenal into a physicalist worldview, the result is that the existence of the physical must be fit into a phenomenal worldview. Surely this inversion is not emblematic of physicalism.

In addition to failing to meet some of the most important motivations for physicalism, subjective physicalism also has some odd consequences that are not typical of physicalist views. The first one is that it results in the impossibility of humans attaining knowledge of the complete physical truths. The second is that postulating irreducibly subjective properties of phenomenal experience fits more naturally with a supernatural ontology than a physicalist ontology.

The first odd consequence of physicalism is that it renders human knowledge of the completed physical sciences to be impossible. Recall that subjective physicalism redefines the physical so that it includes the ontologically irreducible properties of subjective experience. It is certainly odd that on their conception of the physical it requires the person who is doing the ideal physics to have the capacity to have certain subjective experiences. Presumably, both Searle and Howell would admit that someone born blind or deaf is not able to know all the truths of physics because they lack the capacity to have certain subjective experiences. Likewise, since human beings are surely not in a position to know every possible type of subjective conscious experience, they are committed to the position that human beings cannot have a complete physical theory of the world. After all, bats have phenomenal experiences of perception-by-echolocation, and

no human could know what those experiences are like.

Another odd consequence of the ontological commitments of subjective physicalism is that it provides evidence for supernatural theism. Recently philosophers such as Robert Adams, Richard Swinburne, Charles Taliaferro, and J. P. Moreland have argued that the ontological irreducibility of consciousness constitutes *prima facie* evidence in favor of theism and against physicalism.[34] The intuitive idea behind their arguments is that the ontological irreducibility of consciousness fits with the metaphysics of theism where the most fundamental reality is God, a being that is akin to a disembodied, non-physical mind. Furthermore, they claim that an ontology populated with irreducible subjective states of consciousness is not most naturally paired with physicalism, which is primarily motivated by the desire to keep one's ontology within the boundary of the physical sciences. Of course, the point I am making is not that subjective physicalism somehow entails theism. Rather, the point is that the admission of irreducible ontological subjectivity appears to betray the motivations of physicalism, and many philosophers have used this to argue for a non-physical ontology.

In sum, subjective physicalism is not a viable alternative to property dualism in repudiating the Knowledge Argument. First, it fails to fit with the traditional motivations and results of physicalism. The position is thereby

---

[34] Adams (1987); Taliaferro (1994); Swinburne (2004), pp. 192-218; Moreland (2008); Moreland (2009).

physicalist in name only. Second, by failing to meet the typical motivations for physicalism, the actual difference between subjective physicalism and property dualism is very slight, if not altogether undetectable. For these reasons, subjective dualism is not a viable alternative for physicalists who wish to avoid the anti-physicalist implications of the Knowledge Argument.

6.3 Concluding Remarks

Contrary to some critics, the Knowledge Argument does not prove too much by refuting any systematic metaphysical account. The argument succeeds in hitting its target, namely refuting physicalism, while preserving the possibility of dualism's being true. While certain accounts of dualism may not survive the Knowledge Argument, it would be a gross mistake to think all accounts of dualism suffer the same fate. Given that knowledge of the phenomenal requires the subject to be acquainted with those properties, the dualist can plausibly reject that textbook descriptions in black-and-white could provide a systematic knowledge of all the truths of dualism.

Physicalists who attempt to classify the phenomenal properties of conscious experience as part of the physical face severe problems. Their position fails to align with many of the standard motivations and consequences that have characterized physicalist ontologies. In fact, subjective physicalism is virtually indistinguishable from property dualism. While the strategy of "if you can't beat 'em, join 'em" is sometimes the wise one, the result is that the position is

physicalist in name only.

CONCLUSION

The thesis of this dissertation is that our knowledge of the intrinsic nature of conscious experience is incompatible with physicalism's being true. It may be fair to say that this project raises more questions than answers. I have not tried to explain the origin or the kind of reality that accounts for the features of consciousness. Nor have I provided a general theory of the causal relation between the physical and non-physical. My project, I hope, is much more modest. My aim has been to show that physicalism does not have the resources to rebut the Knowledge Argument.

My defense of the Knowledge has followed two approaches. First, I have put forward arguments that directly support the Physical Knowledge Intuition (PKI) and the Knowledge Intuition (KI). For the PKI, I have laid out criteria for providing a robust account of physicalism, and then I surveyed a number of physicalist positions to determine how to understand physicalism. My support for the KI consisted of arguing for a particular approach of epistemology grounded in direct acquaintance and showing how the KI follows from it. My second approach has illustrated intuitive reasons that offer *prima facie* support for both the PKI and KI.

Since most physicalists will likely reject my epistemological commitments, I have strengthened my case by addressing various responses to the *prima facie* case for the Knowledge Argument. By grouping physicalist objections to the Knowledge Argument in three camps, I have canvassed the stock of physicalist

responses to the Knowledge Argument. Through this process, I have provided sufficient reasons to reject these various physicalist responses. Consequently, physicalism does not have the resources to provide a convincing response to the Knowledge Argument.

The final chapter considered a structural question about the Knowledge Argument. Some critics have alleged that the Knowledge Argument proves too much—it proves any systematic objective account of reality cannot account for Mary's new knowledge. The answer to this puzzle is to recognize that the intrinsic nature of conscious experience is a fundamental part of dualism that cannot be known in an analogous way as knowing the fundamental truths of the physical.

While many mysteries about the nature of consciousness remain, what I have shown is that physicalism cannot be sustained given what we do know about phenomenal experience. While this conclusion may be disappointing to some, I believe that it is liberating in a certain way. Having discarded the stringent demands of providing a systematic philosophy within the bounds of physicalism, we are now free to pursue a broader range of possibilities that allow for our world to be open to a reality beyond the strictly physical realm.

BIBLIOGRAPHY

Adams, R. 1987. "Flavors, Colors, and God." In R. Adams's, *The Virtue of Faith and Other Essays in Philosophical Theology* (pp. 243-262). Oxford: Oxford University Press.

Alter, T. 1998. "A Limited Defense of the Knowledge Argument." *Philosophical Studies* 90, pp. 35-56.

Alter, T. 2006. "The Knowledge Argument against Physicalism." In J. Fieser's (ed.), *Internet Encyclopedia of Philosophy*. URL=<http://www.iep.utm.edu/k/know-arg.htm>.

Alter, T. 2007. "Does Representationalism Undermine the Knowledge Argument?" In T. Alter and S. Walter's (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (pp. 65-76). Oxford: Oxford University Press.

Anselm and Gaunilio. 1077-1078. *Proslogion with the Replies of Gaunilo and Anselm*. Many editions and translations.

Armstrong, D. 1968. *A Materialist Theory of Mind*. New York: Routledge.

Audi, R. 2003. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. 2d. ed. New York: Routledge.

Bach-y-Rita, P. 1996. "Sensory Substitution and Qualia." In J. Proust's (ed.), *Perception et Intermodalité* (pp. 81-100). Paris: Presses Universitaires de France; reprinted in Alva Noë and Evan Thompson's (eds.), *Vision and Mind* (pp. 497-514). Cambridge, Mass.: MIT Press, 2002.

Baker, L. R. 1987. *Saving Belief*. Princeton: Princeton University Press.

Baker, L. R. 2000. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.

Bealer, G. and Koons, R. C. 2010. "Introduction." In G. Bealer and R. C. Koons' (eds.) *The Waning of Materialism* (pp. ix-xxxi). Oxford: Oxford University Press.

Beaton, M. 2005. "What RoboDennett Still Doesn't Know." *Journal of Consciousness Studies* 12, no. 1, pp. 3-25.

Beauregard M. and O'Leary, D. 2008. *The Spiritual Brain: A Neuroscientist's Case for the Existence of the Soul*. New York: Harper-Collins.

Berkeley, G. 1710. *Principles of Human Knowledge*. In M. R. Ayers's (ed.), *Philosophical Works: Including the Works on Vision* (pp. 61-127). Totowa, NJ: Rowman and Littlefield, 1975. All references correspond with the standardized section numbers for this work.

Berkeley, G. 1713. *Three Dialogues between Hylas and Philonous*. In M. R. Ayers's (ed.), *Philosophical Works: Including the Works on Vision* (pp. 129-207). Totowa, NJ: Rowman and Littlefield, 1975. All references correspond with the standardized page numbers for this work.

Bickle, J. 2008. "Multiple Realization." In Edward N. Zalta's (ed.) *Stanford Encyclopedia (Fall 2008 Edition)*. URL = <http://plato.stanford.edu/archives/fall2008/entries/multiple-realizability/>.

Bigelow, J. and Pargetter, R. 1990. "Acquaintance with Qualia." *Theoria* 61, pp. 129-147; reprinted in P. Ludlow, Y. Nagasawa, and D. Stoljar's (eds.), *There's Something about Mary* (pp. 179-195). Cambridge, Mass.: MIT Press, 2004.

Block, N. 1978. "Troubles with Functionalism." In C. W. Savage's (ed.), *Perception and Cognition: Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science*, Vol. 9 (pp. 261-325). Minneapolis: University of Minnesota Press.

Block, N. 1990. "Inverted Earth." *Philosophical Perspectives* 4, pp. 53-79.

Boghossian, P. 1990. "The Status of Content." *Philosophical Review* 99, pp. 157-184.

Boghossian, P. 1991. "The Status of Content Revisited." *Pacific Philosophical Quarterly* 71, pp. 264-278.

BonJour, L. 1985. *The Structure of Empirical Knowledge*. Cambridge Mass.: Harvard University Press.

BonJour, L. 2003. "A Version of Internalist Foundationalism." In *Epistemic Justification* (pp. 3-96). Malden, MA.: Blackwell.

BonJour, L. 2005. "What is it Like to be a Human (Instead of a Bat)?" In L. BonJour and A. Baker's (eds.), *Philosophical Problems: An Annotated Anthology* (pp. 397-407). New York: Pearson. Online at: http://faculty.washington.edu/bonjour/Unpublished%20articles/MARTIAN.html.

BonJour, L. 2010. "Against Materialism." In G. Bealer and R. C. Koons's (eds.), *The Waning of Materialism* (pp. 3-23). Oxford: Oxford University Press.

Boyd, R. 2008. "Scientific Realism." In Edward N. Zalta's (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <http://plato.stanford.edu/archives/fall2008/entries/scientific-realism/>.

Braine, D. 1992. *The Human Person: Animal and Spirit*. Notre Dame, IN: University of Notre Dame Press.

Brentano, F. 1874. *Psychology from an Empirical Standpoint*. Translated by T. Rancurello, D. Terrell, and L. McAllister. New York: Routledge Press, 1973.

Broad, C. D. 1925. *The Mind and its Place in Nature*. London: Routledge & Kegan Paul LTD.

Burge, T. 1982. "Other Bodies." In A. Woodfield's (ed.), *Thought and Object: Essays on Intentionality* (pp. 97-120). Oxford: Clarendon Press.

Byrne, A. 2006. "Inverted Qualia." In Edward N. Zalta's (ed.), *The Stanford Encyclopedia of Philosophy (Winter 2008 Edition)*, URL = <http://plato.stanford.edu/archives/win2008/entries/qualia-inverted/>.

Carruthers, P. 2004. "Phenomenal Concepts and Higher-Order Experiences." *Philosophy and Phenomenological Research* 68, pp. 316-336.

Castañeda, H. N. 1967. "Indicators and Quasi-Indicators." *American Philosophical Quarterly* 4, pp. 85–100.

Chalmers, D. 1995a. "Facing up to the Problem of Consciousness." *Journal of Consciousness Studies* 2, pp. 200-219.

Chalmers, D. 1995b. "Absent Qualia, Fading Qualia, Dancing Qualia." In T. Metzinger's (ed.), *Conscious Experience* (pp. 309-328). Exeter: Imprint Academic.

Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.

Chalmers, D. 2003. "Content and Epistemology of Phenomenal Belief." In Q. Smith and A. Jokic's (eds.), *Consciousness: New Philosophical Essays* (pp. 220-272). Oxford, Oxford University Press.

Chalmers, D. 2004. "Phenomenal Concepts and the Knowledge Argument." In P. Ludlow, Y. Nagasawa, and D. Stoljar's (eds.), *There's Something about Mary* (pp. 269-298). Cambridge, Mass.: MIT Press.

Chalmers, D. 2007. "Phenomenal Concepts and the Explanatory Gap. In T. Alter and S. Walter's (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (pp. 167-194). Oxford: Oxford University Press.

Chalmers, D. and Frank J. 2001. "Conceptual Analysis and Reductive Explanation." *Philosophical Review* 110, pp. 315-361.

Chisholm, R. 1967. "Brentano on Descriptive Psychology and the Intentional." In Edward N. Lee and Maurice Mandelbaum's (eds.), *Phenomenology and Existentialism* (pp. 6-23). Baltimore: Johns Hopkins University Press.

Churchill, J. R. and O'Connor, T. 2010. "Nonreductive Physicalism or Emergent Dualism? The Argument from Causation." In G. Bealer and R. C. Koons's (eds.), *The Waning of Materialism* (pp. 261-279.). Oxford: Oxford University Press.

Churchland, P. 1985a. "Reduction, Qualia, and the Direct Introspection of Brain States." *Journal of Philosophy* 82, pp. 8-28.

Churchland, P. 1985b. "Review of Robinson's *Matter and Sense*." *Philosophical Review* 94, pp. 117-120.

Churchland, P. 1988. *Matter and Consciousness*, rev. ed. Cambridge, Mass.: MIT Press.

Churchland, P. 1995. *The Engine of Reason, The Seat of the Soul*. Cambridge, Mass.: MIT Press.

Churchland, P. 2004. "Knowing Qualia: A Reply to Jackson (With Postscript: 1997)." In P. Ludlow, Y. Nagasawa, and D. Stoljar's (eds.), *There's Something about Mary* (pp. 163-178). Cambridge, Mass.: MIT Press.

Churchland, P. S. 1986. *Neurophilosophy: Towards a Unified Science of the Mind-Brain*. Cambridge, Mass.: MIT Press.

Conee, E. 1994. "Phenomenal Knowledge." *Australasian Journal of Philosophy* 72, no. 2, pp. 136-150.

Conee, E. 2005. "The Comforts of Home." *Philosophy and Phenomenological Research* 70, no. 2, pp. 444-451.

Cooney, B. 2000. "Introduction." In B. Cooney's (ed.), *The Place of Mind* (pp. 1-10). Belmont, CA: Wadsworth.

Crane, T. and Mellor, D. H. 1990. "There is No Question of Physicalism." *Mind* 99, pp. 185-206.

Davidson, D. 1970. "Mental Events." In L. Foster and J. W. Swanson's (eds.), *Experience and Theory* (pp. 79-101). Amherst, Mass.: University of Massachusetts Press.

Davidson, D. 1986. "A Coherence Theory of Truth and Knowledge." In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (pp. 423-438). Malden, MA: Blackwell.

Dennett, D. 1971. "Intentional Systems." *The Journal of Philosophy* 68, pp. 87-106.

Dennett, D. 1987. *The Intentional Stance*. Cambridge, Mass.: MIT Press.

Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown, & Co.

Dennett, D. 1996. *Kinds of Mind*. Basic Books: New York.

Dennett, D. 2005. *Sweet Dreams*. Cambridge, Mass.: MIT Press.

Dennett, D. 2007. "What RoboMary Knows." In Torin Alter and Sven Walter's (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (pp. 15-31). Oxford University Press.

Descartes, R. 1641. *Meditations on First Philosophy*. Translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch in *The Philosophical Writings of Descartes: Volume II* (pp. 1-62). Cambridge: Cambridge University Press, 1984.

Dilley, F. B. 2004. "Taking Consciousness Seriously: A Defense of Cartesian Dualism." *International Journal for Philosophy of Religion* 55, pp. 135-153.

Donnellan, K. S. 1966. "Reference and Definite Descriptions." *Philosophical Review* 77, pp. 281-304.

Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, Mass.: MIT Press.

Dretske, F. 1996. "Phenomenal Externalism or If Meanings Ain't in the Head, Where are the Qualia?" In V. Villanueva's (ed.), *Philosophical Issues 7: Perception* (pp. 143-158). Atascadero, CA: Ridgeview Publishing.

Endicott, R. P. 1995. "The Refutation by Analogous Ectoqualia." *The Southern Journal of Philosophy* 33, pp. 19-30.

Evans, G. 1973. "The Causal Theory of Names." *Proceedings of the Aristotelian Society,* supplementary volume, 47, pp. 187-208.

Fales, E. 1990. *Causation and Universals*. New York: Routledge.

Fales, E. 1996. *A Defense of the Given*. Lanham, MD: Rowman and Littlefield.

Feldman, R. 2004. "The Justification of Introspective Beliefs." In E. Conee and R. Feldman's, *Evidentialism: Essays in Epistemology* (pp. 199-218). Oxford: Oxford University Press.

Feyerabend, P. K. 1958. "An Attempt at a Realistic Interpretation of Experience." *Proceedings of the Aristotelian Society* 58, pp. 143-170.

Fodor, J. 1974. "Special Sciences, or The Disunity of Science as a Working Hypothesis." *Synthese* 28, pp. 97-115.

Foster, J. 1991. *Immaterial Self*. New York: Routledge.

Fumerton, R. 1988. "Russelling Causal Theories of Reference." In C. W. Savage's (ed.), *Rereading Russell: Essays on Bertrand Russell's Metaphysics and Epistemology* (pp. 108-118). Minneapolis: University of Minnesota Press.

Fumerton, R. 1995. *Metaepistemology and Skepticism*. Lanham, MD: Rowman and Littlefield.

Fumerton, R. 1998. "Knowledge by acquaintance and description." In E. Craig's (ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge. Retrieved May 13, 2009, from http://www.rep.routledge.com.proxy.lib.uiowa.edu/article/P003SECT1

Fumerton, R. 2001. "Classical Foundationalism." In M. R. DePaul's (ed.), *Resurrecting Old-Fashioned Foundationalism* (pp. 3-20). Lanham, MD: Rowman and Littlefield.

Fumerton, R. 2005. "Speckled Hens and Objections of Acquaintance." *Philosophical Perspectives* 19, pp. 121-138.

Fumerton, R. 2006a. *Epistemology*. Malden, MA: Blackwell.

Fumerton, R. 2006b. "Direct Realism, Introspection, and Cognitive Science." *Philosophy and Phenomenological Research* 73, no. 3, pp. 680-695.

Fumerton, R. 2008. "Knowledge by Acquaintance vs. Description." In Edward N. Zalta's (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <http://plato.stanford.edu/archives/fall2008/entries/knowledge-acquaindescrip/>.

Fumerton, R. 2009. "Luminous Enough for a Cognitive Home." *Philosophical Studies* 142, no. 1, pp. 67-76.

Gertler, B. 1999. "A Defense of the Knowledge Argument." *Philosophical Studies* 93, pp. 317-336.

Goetz, S. 2005. "Substance Dualism." In J. B. Green and S. L. Palmer's (eds.), *In Search of the Soul: Four Views of the Mind-Body Priblem* (pp. 33-60). Downers Grove, Il.: Inter-Varsity Press.

Griffin, D. 1998. *Unsnarling the World-Knot: Consciousness, Freedom and the Mind-Body Problem*. Berkeley: University of California Press.

Grice, H. P. 1989. *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press.

Harman, G. 1990. "The Intrinsic Quality of Experience." In J. Tomberlin's (ed.), *Philosophical Perspectives 4: Action Theory and Philosophy of Mind* (pp. 31-52). Atascadero, CA: Ridgeview Publishing Company.

Hart, W. D. 1988. *Engines of the Soul*. Cambridge: Cambridge University Press.

Hasker, W. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.

Haugeland, J. 1982. "Weak Supervenience." *American Philosophical Quarterly* 19: 93-104.

Haugeland, J. 1984. "Ontological Supervenience." *The Southern Journal of Philosophy* 22, supplement pp. 1-12.

Hawthorne, J. 2005. "Knowledge and Evidence." *Philosophy and Phenomenological Research* 70, no. 2, pp. 452-458.

Hempel, C. 1970. "Reduction: Ontological and Linguistic Facets." In S. Morgenbesser, et al., (eds.), *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel* (pp. 179-199). New York: St. Martin's Press.

Hobbes, T. 1660. *The Leviathan*. Translated by E. Curley. Indianapolis: Hackett Publishing, 1994.

Holbach, Paul-Henri Thiry. 1770. *System of Nature*. Translated by H. D. Robinson. New York: Burt Franklin, 1970.

Horgan, T and Tienson, J. "The Intentionality of Phenomenology and the Phenomenology of Intentionality." In D. Chalmers's (ed.), *The Philosophy of Mind: Classical and Contemporary Readings* (pp. 520-533). Oxford: Oxford University Press.

Howell, R. J. 2007. "The Knowledge Argument and Objectivity." *Philosophical Studies* 135, pp. 145-177.

Howell, R. J. 2009. "The Ontology of Subjective Physicalism." *Noûs* 43, no. 2, pp. 315-345.

Hume, D. 1748. *An Enquiry concerning Human Understanding*; reprinted, in T. Beauchamp's (ed.), *An Enquiry concerning Human Understanding*. Oxford: Oxford University Press, 1999.

Ismael, J. 1999. "Science and the Phenomenal." *Philosophy of Science* 66, pp. 351-369.

Jackson, F. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32, pp. 126-136.

Jackson, F. 1986. "What Mary Didn't Know." *The Journal of Philosophy* 83, pp. 291-295.

Jackson, F. 1995. "Postscript on 'What Mary didn't know.'" In P. Moser and J. Trout, (eds.), *Contemporary Materialism* (pp. 184-189). New York: Routledge; reprinted in ; reprinted in P. Ludlow, Y. Nagasawa, and D. Stoljar's (eds.), *There's Something about Mary* (pp. 409-15). Cambridge, Mass.: MIT Press, 2004.

Jackson, F. 1998. *From Metaphysics to Ethics*. Oxford: Oxford University Press.

Jackson, F. 2002. "Mind and Illusion." Paper presented at the Royal Institute of Philosophy; reprinted in P. Ludlow, Y. Nagasawa, and D. Stoljar's (eds.), *There's Something about Mary* (pp. 421-442). Cambridge, Mass.: MIT Press, 2004.

Jackson, F. 2007a. "A Priori Physicalism." In B. McLaughlin and J. Cohen's (eds.), *Contemporary Debates in Philosophy of Mind* (pp. 185-199). Malden, Mass.: Blackwell.

Jackson, F. 2007b. "The Knowledge Argument, Diaphanousness, Representationalism." In Torin Alter and Sven Walter's (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (pp. 52-64). Oxford University Press.

Jacquette, D. 1995. "The Blue Banana Trick: Dennett on Jackson's Color Scientist." *Theoria* 61, pp. 217-230.

Kim, J. 1989. "The Myth of Non-Reductive Physicalism." *Proceedings and Addresses of the American Philosophical Association* 63, no. 3, pp. 31-47.

Kim, J. 1992a. "Multiple Realization and the Metaphysics of Reduction." *Philosophy and Phenomenological Research* 52, pp. 1-26.

Kim, J. 1992b. " 'Downward Causation' in Emergentism and Nonreductive Materialism." In A. Berckermann, H. Flohr, and J. Kim's (eds.), *Emergence or Reduction?* (pp. 119-138). Berlin: De Gruyter.

Kim, J., ed. 1993a. *Supervenience and Mind*. Cambridge: Cambridge University Press.

Kim, J. 1993b. " 'Strong' and 'Global' Supervenience Revisited." In J. Kim's, *Supervenience and Mind* (pp. 79-91). Cambridge: Cambridge University Press.

Kim, J. 1997. "The Mind-Body Problem: Taking Stock after 40 Years." *Philosophical Perspectives* 11, pp. 185-207.

Kim, J. 1998. *Mind in a Physical World*. Cambridge, Mass.: MIT Press.

Kim, J. 1999. "Making Sense of Emergence." *Philosophical Studies* 95, pp. 3-36.

Kim, J. 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.

Kornblith, H. 2002. *Knowledge and its Place in Nature*. Oxford: Oxford University Press.

Kripke, S. 1979. "A Puzzle about Belief." In A. Margalit's (ed.), *Meaning and Use* (pp. 239-283). Dodrecht: D. Reidel.

Kripke, S. 1980. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.

Kuhn, T. S. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Leftow, B. 2010. "Soul, Mind, and Brain." In G. Bealer and R. C. Koons's (eds.), *The Waning of Materialism* (pp. 395-415). Oxford: Oxford University Press.

Leibniz, G. 1714. *Monadology*. Translated by Robert Latta. Translation originally published 1898; reprinted Charleston, SC: Forgotten Books, 2008.

Levin, J. 2007. "What is a Phenomenal Concept?" In T. Alter and S. Walter's (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (pp. 87-110). Oxford: Oxford University Press.

Lewis, C. I. 1929. *Mind and the World-Order*. New York: Charles Scribner's Sons.

Lewis, D. 1966. "An Argument for the Identity Theory." *Journal of Philosophy* 63, pp. 17-25.

Lewis, D. 1988. "What Experiences Teaches." *Proceedings of the Russellian Society* 13, pp. 29-57; reprinted in P. Ludlow, Y. Nagasawa, and D. Stoljar's (eds.), *There's Something about Mary* (pp. 77-103). Cambridge, Mass.: MIT Press, 2004.

Lewis, D. 1992. "Multiple Realization and the Metaphysics of Reduction." *Philosophy and Phenomenological Research* 52, pp. 309-335.

Lewis, D. 1994. "Reduction of Mind." In S. Guttenplan's (ed), *A Companion to the Philosophy of Mind* (pp. 412-431). Oxford: Blackwell.

Loar, B. 1997. "Phenomenal States (Revised Version)." In N. Block, O. Flanagan, and G. Guzeldere's (eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 597-608). Cambridge, Mass.: MIT Press; reprinted in P. Ludlow, Y. Nagasawa, and D. Stoljar's (eds.), *There's Something about Mary* (pp. 219-239). Cambridge, Mass.: MIT Press.

Locke, J. 1689. *An Essay Concerning Human Understanding*, ed. P. Phemister. Oxford: Oxford University Press, 2008.

Lockwood, M. 1989. *Mind, Brain, and the Quantum.* Oxford: Oxford University Press.

Lowe, E. J. 2010. "Substance Dualism: A Non-Cartesian Approach." In G. Bealer and R. C. Koons's (eds.), *The Waning of Materialism* (pp. 439-561). Oxford: Oxford University Press.

Ludlow, P. 2009. "Descriptions." In Edward N. Zalta's (ed.), *The Stanford Encyclopedia of Philosophy (Spring 2009 Edition)*, URL = <http://plato.stanford.edu/archives/spr2009/entries/descriptions/>.

Lycan, W. 1996. *Consciousness and Experience.* Cambridge, Mass.: MIT Press.

Lycan, W. 2001. "The Case for Phenomenal Externalism." In J. E. Tomberlin's (ed.), *Philosophical Perspectives 15: Metaphysics* (pp. 17-35). Atascadero, CA: Ridgeview Publishing.

Lycan, W. 2008. "Representational Theories of Consciousness." In Edward N. Zalta's (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <http://plato.stanford.edu/archives/fall2008/entries/consciousness-representational/>.

Maddell, G. 1988. *Mind and Materialism*. Edinburgh: Edinburgh University Press.

Markie, P. 2009. "Classical Foundationalism and Speckled Hens." *Philosophy and Phenomenological Research* 79, no. 1, pp. 190-206.

Maxwell, G. 1978. "Rigid Designators and Mind-Brain Identity." In C. W. Savage's (ed.), *Perception and Cognition: Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science*, Vol. 9 (pp. 365-403). Minneapolis: University of Minnesota Press.

McConnell, J. 1994. "In Defense of the Knowledge Argument." *Philosophical Topics* 22, pp. 157-187.

McLaughlin, B. and Bennett, K. 2008. "Supervenience." In Edward N. Zalta's (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <http://plato.stanford.edu/archives/fall2008/entries/supervenience/>.

McGinn, C. 1999. *The Mysterious Flame: Conscious Minds in a Material World*. New York: Basic Books.

McGeer, V. 2003. "The Trouble with Mary." *Pacific Philosophical Quarterly* 84, no. 4, pp. 384-393.

McGrew, T. 1995. *The Foundations of Knowledge*. Lanham, MD: Littlefield Adams.

McGrew, T. and McGrew, L. 2007. *Internalism and Epistemology*. New York: Routledge.

McMullen, C. 1985. " 'Knowing what it's Like' and the Essential Indexical." *Philosophical Studies* 48, pp. 211-233.

Melnyk, A. 2003. *A Physicalist Manifesto: A Thoroughly Modern Materialism*. Cambridge: Cambridge University Press.

Menuge, A. 2004. *Agents Under Fire: Materialism and the Rationality of Science*. Lanham, MD: Rowman and Littlefield.

Merricks, T. 2003. *Objects and Persons*. Oxford: Oxford University Press.

Moore, G. E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.

Moore, G. E. 1922. "The Refutation of Idealism." In G. E. Moore's, *Philosophical Studies* (pp. 1-20). London: Kegan Paul.

Moreland, J. P. 2003. "The Knowledge Argument Revisited." *International Philosophical Quarterly* 43, no. 2, pp. 219-228.

Moreland, J. P. 2008. *Consciousness and the Existence of God*. New York: Routledge.

Moreland, J. P. 2009. "The Argument from Consciousness." In W. L. Craig and J. P. Moreland's (eds.), *The Blackwell Companion to Natural Theology* (pp. 282-343). Malden, Mass.: Blackwell.

Moser, P. K. 1985. *Empirical Justification*. Dordrecht: D. Reidel.

Moser, P. K. 1989. *Knowledge and Evidence*. Cambridge: Cambridge University Press.

Nagasawa, Y. 2002. "The Knowledge Argument against Dualism." *Theoria* 68, pp. 205-223.

Nagasawa, Y. and Stoljar, D. 2004. "Introduction." In P. Ludlow, Y. Nagasawa, and D. Stoljar's (eds.), *There's Something about Mary* (pp. 1-36). Cambridge, Mass.: MIT Press.

Nagel, E. 1961. *The Structure of Science*. New York: Harcourt, Brace.

Nagel, T. 1974. "What is it Like to be a Bat?" *Philosophical Review* 83, pp. 435-450.

Nagel, T. 1979. "Panpsychism." In *Mortal Questions* (pp. 181-195). Cambridge, Eng.: Cambridge University Press.

Nagel, T. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

Nagel, T. 1998. "Conceiving the Impossible and the Mind-Body Problem." *Philosophy* 73, pp. 337-352.

Neale, S. 1990. *Descriptions*. Cambridge, Mass.: MIT Press.

Nemirow, L. 1990. "Physicalism and the Cognitive Role of Acquaintance." In W. Lycan's (ed.), *Mind and Cognition: A Reader* (pp. 490-499). Oxford: Blackwell.

Nemirow, L. 2007. "So This is What It's Like: A Defense of the Ability Hypothesis." In Torin Alter and Sven Walter's (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (pp. 32-51). Oxford University Press.

Nida-Rümelin, M. 1995. "What Mary Couldn't Know: Belief about Phenomenal States." In T. Metzinger's (ed.), *Conscious Experience* (pp. 219-241). Exeter: Imprint Academic. Reprinted in P. Ludlow, Y. Nagasawa, and D. Stoljar's (eds.), *There's Something about Mary* (pp. 241-265). Cambridge, Mass.: MIT Press, 2004.

Nida-Rümelin, M. 1996. "Pseudonormal Vision: An Actual Case of Qualia Iversion?" *Philosophical Studies* 82, pp. 145-157.

Nida-Rümelin, M. 2007. "Grasping Phenomenal Properties." In T. Alter and S. Walter's (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (pp. 307-338). Oxford: Oxford University Press.

Nida-Rümelin, M. 2008. "Qualia: The Knowledge Argument." In Edward N. Zalta's (ed.) *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, URL = <http://plato.stanford.edu/archives/fall2008/entries/qualia-knowledge/>.

O'Dea, J. 2002. "The Indexical Nature of Sensory Concepts." *Philosophical Papers* 31, pp. 169-181.

Papineau, D. 2002. *Thinking about Consciousness*. Oxford: Oxford University Press.

Papineau, D. 2007. "Phenomenal and Perceptual Concepts." In Torin Alter and Sven Walter's (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (pp. 111-144). Oxford University Press.

Penfield, W. 1975. *The Mystery of the Mind*. Princeton: Princeton University Press.

Pereboom, D. 2002. "Robust Nonreductive Materialism." *Journal of Philosophy* 99, pp. 499-531.

Perry, J. 1979. "The Problem of the Essential Indexical." *Noûs* 13, no. 1, pp. 3-21.

Perry, J. 2001. *Knowledge, Possibility, and Consciousness*. Cambridge, Mass.: MIT Press.

Pitt, D. 2004. "The Phenomenology of Cognition or *What Is It Like to Think That P*?" *Philosophy and Phenomenological Research* 69, no. 1, pp. 1-36.

Plantinga A. 2006. "Against Materialism." *Faith and Philosophy* 23, no. 1, pp. 3-32.

Place, U. T. 1956. "Is Consciousness a Brain Process?" *British Journal of Psychology* 47, pp. 44-50.

Poland, J. 1994. *Physicalism: The Philosophical Foundations*. Oxford: Oxford University Press.

Popper, K and Eccles, J. 1977. *The Self and its Brain*. New York: Routledge.

Poston, T. 2007. "Acquaintance and the Problem of the Speckled Hen." *Philosophical Studies* 132, pp. 331-346.

Price, H. H. 1950. *Perception*. 2d. ed. London: Meuthen.

Psillos, S. 1999. *Scientific Realism: How Science Tracks the Truth*. New York: Routledge.

Putnam, H. 1967. "Psychological Predicates." In W. H. Capitan and D. D. Merrill's (eds.), *Art,, Mind and Religion* (pp. 37-48). Pittsburgh: University of Pittsburgh Press.

Putnam, H. 1975a. "Brains and Behavior." In *Philosophical Papers: Volume 2 Mind, Language and Reality* (pp. 325-341). Cambridge: Cambridge University Press.

Putnam, H. 1975b. "The Meaning of 'Meaning.'" *Minnesota Studies in the Philosophy of Science* 7, pp. 131-193.

Putnam, H. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.

Putnam, H. 1996. "Introduction." In A. Pessin's and S. Goldberg's (eds.), *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "The Meaning of 'Meaning'"* (pp. xv-xxii). Armonk, NY: M. E. Sharpe.

Quine, W. V. O. 1956. "Quantifiers and Propositional Attitudes." *The Journal of Philosophy* 53, pp. 177-187.

Quinn, P. 1997. "Tiny Selves: Chisholm on the Simplicity of the Soul." In L. E. Hahn's (ed.), *The Philosophy of Roderick M. Chisholm* (pp. 55-67). Peru, IL: Open Court.

Reed, B. 2006. "Shelter for the Cognitively Homeless." *Synthese* 148, no. 2, pp. 303-308.

Reimer, M. 1992. "Incomplete Descriptions." *Erkenntnis* 37, pp. 347-363.

Reppert, V. 1992. "Eliminative Materialism, Cognitive Suicide, and Begging the Question." *Metaphilosophy* 23, pp. 378-392.

Rey, G. 1998. " A Narrow Representationalist Account of Qualitative Experience." In J. E. Tomberlin's (ed.), *Philosophical Perspectives 12: Language, Mind, and Ontology* (pp. 435-457). Atascadero, CA.: Ridgeview Publishing.

Robinson, H. M. 1982. *Matter and Sense*. Cambridge: Cambridge University Press.

Robinson, H. M. 1993a. "Dennett on the Knowledge Argument." *Analysis* 53, no. 3, pp. 174-177; reprinted in P. Ludlow, Y. Nagasawa, and D. Stoljar's (eds.), *There's Something about Mary* (pp. 69-73). Cambridge, Mass.: MIT Press, 2004.

Robinson, H. M. 1993b. "The Anti-Materialist Strategy and the 'Knowledge Argument.' " In H. M. Robinson's (ed.), *Objections to Physicalism* (pp. 159-183). Oxford: Oxford University Press.

Robinson, H. M. 2008. "Dualism." In Edward N. Zalta's (ed.), *The Stanford Encyclopedia (Fall 2008) Edition*, URL = <http://plato.stanford.edu/archives/fall2008/entries/dualism/>.

Rorty, R. 1965. "Mind-Body Identity, Privacy, and Categories." *The Review of Metaphysics* 19, no.1, pp. 15-35.

Rorty, R. 1970. "In Defense of Eliminative Materialism." *The Review of Metaphysics* 24, no. 1, pp. 112-121.

Rosenberg, G. 2005. *A Place for Consciousness: Probing the Deep Structure of the Natural World*. Oxford: Oxford University Press.

Russell, B. 1905. "On Denoting." *Mind* 14, pp. 479-493.

Russell, B. 1910. "Knowledge by Acquaintance and Knowledge by Description." *Proceedings of the Aristotelian Society* 11, pp. 108-128.

Russell, B. 1912. *Problems of Philosophy*, ed. John Perry. Oxford: Oxford University Press, 1997.

Russell, B. 1921. *The Analysis of Mind*. New York: MacMillan.

Russell, B. 1927. *The Analysis of Matter*. London: Kegan Paul.

Russell, B. 1957. "Mr. Strawson on Referring." *Mind* 66, pp. 385-389.

Ryle, G. 1949. *The Concept of Mind.* Chicago: University of Chicago Press.

Schiffer, S. 2005. "Russell's Theory of Definite Descriptions." *Mind* 114, pp. 1135-1183.

Searle, J. R. 1983. *Intentionality*. Cambridge, Eng.: Cambridge University Press.

Searle, J. R. 1992. *The Rediscovery of the Mind.* Cambridge, Mass.: MIT Press.

Searle, J. R. 2002. *Consciousness and Language*. Cambridge, Eng.: Cambridge University Press.

Searle, J. R. 2004. *Mind: A Brief Introduction*. Oxford: Oxford University Press.

Sellars, W. 1956. "Empiricism and the Philosophy of Mind." In Herbert Feigl and Michael Scriven's (eds.), *Minnesota Studies in the Philosophy of Science I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis* (pp. 253-329). University of Minnesota Press, Minneapolis.

Shoemaker, S. 2007. *Physical Realization*. Oxford: Oxford University Press.

Skrbina, D. 2005. *Panpsychism in the West.* Cambridge, Mass.: MIT Press.

Smart, J. J. C. 1959. "Sensations and Brain Processes." *Philosophical Review* 68, pp. 141-156.

Sosa, E. 2003a. "Privileged Access." In Q. Smith and A. Jokic's (eds.), *Consciousness: New Philosophical Perspectives* (pp. 273-292). Oxford: Oxford University Press.

Sosa, E. 2003b. "Beyond Internal Foundations to External Virtues." In *Epistemic Justification* (pp. 99-170). Malden, MA: Blackwell.

Sprigge, T. 1983. *A Vindication of Absolute Idealism*. London: Routledge.

Stalnaker, R. C. 2008. *Our Knowledge of the Internal World*. Oxford: Oxford University Press.

Stich, S. 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, Mass: MIT Press.

Stoljar, D. 2001. "Two Conceptions of the Physical." *Philosophy and Phenomenological Research* 62, pp. 253-281.

Stoljar, D. 2005. "Physicalism and Phenomenal Concepts." *Mind & Language* 20, no. 5, pp. 469-494.

Strawson, P. F. 1950. "On Referring." *Mind* 59, pp. 320-334.

Stump, E. 1995. "Non-Cartesian Substance Dualism and Materialism without Reductionism." *Faith and Philosophy* 12, pp. 505-531.

Swinburne, R. 1984. "Personal Identity: The Dualist Theory." In S. Shoemaker and R. Swinburne's, *Personal Identity* (pp. 1-66). Oxford: Blackwell.

Swinburne, R. 1996. *Is there a God?* Oxford: Oxford University Press.

Swinburne, R. 1997. *The Evolution of the Soul*, rev. ed. Oxford: Oxford University Press.

Swinburne, R. 2001. *Epistemic Justification*. Oxford: Oxford University Press.

Swinburne, R. 2004. *The Existence of God*. 2d. ed. Oxford: Oxford University Press.

Swinburne, R. 2009. "Substance Dualism." *Faith and Philosophy* 26, no. 5, pp. 501-513.

Szabó, Z. G. 2000. "Descriptions and Uniqueness." *Philosophical Studies* 101, pp. 29-57.

Taliaferro, C. 1994. *Consciousness and the Mind of God*. Cambridge, Eng.: Cambridge University Press.

Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, Mass.: MIT Press.

Tye, M. 2000. *Consciousness, Color, and Content*. Cambridge, Mass.: MIT Press.

Tye, M. 2003. "A Theory of Phenomenal Concepts." In A. O'Hear's (ed.), *Minds and Persons* (pp. 91-105). Cambridge: Cambridge University Press.

Tye, M. 2009. *Consciousness Revisited: Materialism without Phenomenal Concepts*. Cambridge, Mass.: MIT Press.

Unger, P. 2006. *All the Power in the World*. Oxford: Oxford University Press.

van Gulick, R. 2004. "So Many Ways of Saying No to Mary." In P. Ludlow, Y. Nagasawa, and D. Stoljar's (eds.), *There's Something about Mary* (pp. 365-405). Cambridge, Mass.: MIT Press.

von Fintel, K. 2004. "Would You Believe It? The Present King of France is Back! (Presuppositions and Truth-Value Intuitions)." In M. Reimer and A. Bezuidenthout's (eds.), *Descriptions and Beyond* (pp. 315-341). Oxford: Oxford University Press.

Weiskrantz, L. 2009. *Blindsight: A Case Study Spanning 35 Years*. 2d. ed. Oxford: Oxford University Press.

Wettstein, H. K. 1981. "Demonstrative Reference and Definite Descriptions." *Philosophical Studies* 40, no. 2, pp. 241-257.

White, S. L. 2007. "Property Dualism, Phenomenal Concepts, and the Semantic Premise." In T. Alter and S. Walter's (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press.

White, S. L. 2010. "The Property Dualism Argument." In G. Bealer and R. Koons's (eds.), *The Waning of Materialism* (pp. 89-113). Oxford: Oxford University Press.

Whitehead, A. N. 1929. *Process and Reality: An Essay in Cosmology*. New York: MacMillan.

Williams, M. 1999. *Groundless Belief: An Essay on the Possibility of Epistemology*. 2d. ed. Princeton, NJ.: Princeton University Press.

Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

Wilson, E. O. 1998. *Consilience: The Unity of Knowledge*. New York: Alfred A. Knopf.

Yandell, K. A. 1995. "A Defense of Dualism." *Faith and Philosophy* 12, pp. 548-566.

Zimmerman, D. 2003. "Christians Should Affirm Mind-Body Dualism." In Michael Peterson and Raymond Vanarragon's (eds.), *Contemporary Debates in Philosophy of Religion* (pp. 314-324). Oxford: Blackwell.

Zimmerman, D. 2006. "Dualism in the Philosophy of Mind." In Donald Borchert's (ed.) 2nd ed. *The Encyclopedia of Philosophy* (pp. 113-122). New York: MacMillan.