PHYSICALISM, DUALISM AND THE MIND-BODY PROBLEM

A Dissertation

Submitted to the Graduate School

of the University of Notre Dame

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

by

Dolores G. Morris

_____

Alvin Plantinga, Director

Graduate Program in Philosophy

Notre Dame, Indiana

December 2010

PHYSICALISM, DUALISM AND THE MIND-BODY PROBLEM

Abstract

by

Dolores G. Morris

In this dissertation, I examine the implications of the problem of mental causation and what David Chalmers has dubbed the "hard problem of consciousness" for competing accounts of the mind. I begin, in Chapter One, with a critical analysis of Jaegwon Kim's *Physicalism, or Something Near Enough*. (2005) There, I maintain that Kim's ontology cannot adequately address both the problem of mental causation and the "hard problem of consciousness." In Chapter Two, I examine the causal pairing problem for substance dualism. I demonstrate both that the substance dualist can respond to the argument at no great cost, and that the pairing problem applies, with equal force, to the irreducible qualia posited on Kim's account.

Chapters Three and Four are devoted to what I take to be the central argument against any kind of dualism: the causal exclusion argument. In Chapter Three, I examine dualistic responses to the exclusion argument that grant the causal closure of the physical world. I note that these responses, though technically adequate, are nevertheless theoretically unpalatable. In addition to requiring the dualist to adopt unconventional

attitudes towards causation, responses of this variety also have the unfortunate result of rendering libertarian freedom impossible. Finally, in Chapter Four, I turn my attention to the question of causal closure. I maintain that the causal closure of the physical world, though widely affirmed, is nevertheless extraordinarily difficult to support. In light of Hempel's Dilemma, causal closure is either false, compatible with dualistic interaction, or unacceptably stipulative. There is, I maintain *no* causal closure principle up to the tasks required by the causal exclusion argument. For that reason, I conclude that the dualist ought *not to worry* about causal closure.

CONTENTS

CHAPTER ONE:

PHYSICALISM AND THE MIND-BODY PROBLEM:
A CRITICAL ANALYSIS OF JAEGWON KIM'S FUNCTIONAL REDUCTION

In *Physicalism, or Something Near Enough*, Jaegwon Kim presents an account of the mind according to which most, but not all, mental properties can be reduced to physical ones via functional definitions.[1] While the position at which he arrives fails, ultimately, to be a wholly physicalist account, he nevertheless maintains that it is the physicalist's best bet when it come to resolving the mind-body problem. Furthermore, because Kim also argues for a rejection of substance dualism, he maintains that this position is really *anyone's* best chance at resolving the mind body problem. In what follows, I wish to examine the first of these claims.

I will begin in §1 with a statement of the mind-body problem, though as we shall see the "problem" turns out to be, instead, a collection of related problems. In §2, I will present Kim's treatment and eventual rejection of a series of physicalist responses to the mind-body problem(s). I will offer a reconstruction of Kim's positive account, physicalist reduction via functional definitions, in §3. The remainder of this chapter, §4-8, will

[1] Jaegwon Kim *Physicalism or Something Near Enough*. (Princeton: Princeton University Press, 2005)

consist of a critical analysis of Kim's position. In §4 I will consider a series of questions that are, as of yet, unanswered on Kim's account. I suggest that the answers to these questions will be critically important to an overall assessment of Kim's position. In §5 I pose one additional question: namely, whether or not irreducible qualia can be said to supervene on physical properties given Kim's ontology. §6 and §7 will be devoted to assessing the success of Kim's account in treating both aspects of the mind-body problem for physicalism: the problem of mental causation, and the problem of consciousness. I will conclude in §8.

Physicalism, or Something Near Enough presents a carefully articulated and boldly argued account of the mind. The arguments are impressive, both in scope and in clarity, and ought not to be treated lightly. Nevertheless, I wish to dispute Kim's claim that functional reductions are the key to resolving the mind-body problem. I will argue, instead, that both the problem of mental causation and the problem of consciousness remain problematic given a functionally reductive account. Ultimately, I will suggest that a person—physicalist or otherwise—who is interested in responding to the mind-body problem ought to seek alternative means.

§1. The Mind Body Problem(s)

Kim begins the first chapter of *Physicalism, or Something Near Enough* by noting that, strictly speaking, there is no single "mind-body problem." Instead, the so-called "mind-body problem" is best understood as:

> A cluster of connected problems about the relationship between mind and matter. What these problems are depends on a broader framework of philosophical and

> scientific assumptions and presumptions within which the questions are posed and possible answer are formulated. (7)

A substance dualist, for example, affirms the immateriality of the mind. Most substance dualists are also interactionist dualists, which is to say they affirm two-way causal interaction between the mind and the body. For an interactionist dualist, then, the principle mind-body problem is the problem of accounting for causal interaction between an immaterial substance and a material one. Absent some account of how an immaterial substance could act causally upon a material one, the substance dualist will face difficulties in attempting to explain mental causation.

For a contemporary physicalist, however, there is no such problem. A physicalist need not explain the possibility of causal interaction between "diverse substances" because the physicalist affirms a monism of substance. This is not to say that there is *no* mind-body problem pertinent to physicalism, only that the nature of the problem (or problems, as it were) depends upon the broader ontological context in which it is raised. According to Kim, the mind-body problem for a contemporary physicalist is comprised of two separate, but related, issues: the problem of mental causation, and the problem of consciousness.[2]

The problem of mental causation, as formulated against a physicalist, can take a variety of forms. Broadly stated, the question is this: "How can the mind exercise its causal powers in a causally closed physical world?" (13) Because "the mind," in this context, no longer refers to a distinct type of substance, this question is more often framed in terms of mental *properties* or mental *events* than it is in terms of "the mind" as such.

[2] Again, it seems unlikely that there is only *one* "problem of mental causation." Still, there is *a* problem of mental causation here, and a distinct but related problem of consciousness.

Given that the physical domain is causally closed, how can mental properties exert causal influence in the physical world? How can a mental event be the cause of some physical event? How can *mentality*, however construed, be causally relevant? This is the first component of the physicalist's mind-body problem.

The second component, the problem of consciousness, centers upon the mere presence of mentality in the physical world. As Kim writes, "Why is there, and how can there be, such a thing as the mind, or consciousness, in a physical world?" (13) What Kim calls the "problem of consciousness" is essentially what Chalmers has dubbed "the hard problem of consciousness."[3] In *The Conscious Mind*, Chalmers offers the following illustration of the hard problem of consciousness:

> When someone strikes middle C on the piano, a complex chain of events is set into place. Sound vibrates in the air and a wave travels to my ear. The wave is processed and analyzed into frequencies inside the ear, and a signal is sent to the auditory cortex…All this is not so hard to understand in principle. But why should this be accompanied by an *experience*? And why, in particular, should it be accompanied by *that* experience, with its characteristic rich tone and timbre?[4]

This problem, itself twofold, is the problem of consciousness. Why is there phenomenal experience at all? Why does phenomenal experience occur in the precise way that it does, rather than in some other way?

For contemporary *reductive* physicalists—physicalists who believe that all mental properties or states can be reduced to physical ones—the problem of consciousness poses two distinct challenges. The first is the task of closing the "explanatory gap" between

---

[3] David I. Chalmers, *The Conscious Mind*: In Search of a Fundamental Theory. (New York & Oxford: Oxford University Press, 1996) (p4 and elsewhere) More specifically, Kim writes that the problem of the explanatory gap is "what David Chalmers has called the 'hard problem' of consciousness." (Kim 2005, p.93-94) Kim's "problem of consciousness" seems also to include the predictive, or epistemic, gap. Still, the basic difficulty is the same: how are we to reconcile phenomenal consciousness with a causally closed, physical world?

[4] (Chalmers 1996, p.5)

phenomenal consciousness and the physical world.[5] This is essentially the problem which we have been discussing; it asks *why* consciousness exists, at all and as it does.[6]

The second challenge is centered upon what Kim calls the *predictive* (or *epistemic*) gap between consciousness and the physical world. Kim writes

> As the emergentists claimed, it seems possible for us to know all about the physiology of a creature, say Thomas Nagel's famously inscrutable bats,[7] but have no idea of the qualitative character of its inner experience. (94)

If reductive physicalism is true, then complete knowledge of the physical properties of a creature should suffice for knowledge of the mental properties of that creature. Given the fundamental physical features of a bat, we ought to be able to conclude the nature of any mental features that that bat might have. Yet this is not the case. Instead, any predictions that we make about the conscious states of animals are based largely upon observed correlations between physical states and conscious ones; in addition to the physical evidence, we rely upon observed *phenomenal* evidence as well. An adequate reductive physicalist account of the mind, then, should be able to close this gap. It should give us the resources to ascribe mental properties by appealing only to a purely physical base domain.

---

[5] This phrase was coined by Joseph Levine in "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly* 64 (1983): 354-361 (Cited in Kim 2005, p.93)

[6] Henceforth, with Chalmers, I will use the term "consciousness" to refer to *phenomenal* consciousness—what Kim and others call "qualia." One tends to think primarily about colors, sounds and the like when one hears "qualia." I maintain, with Chalmers, that phenomenal consciousness is far more widespread than this would indicate; there are qualitative *aspects* to nearly all of our mental states. As such, I prefer the term "consciousness" and, unless otherwise stated, will use it to refer only to the phenomenal aspects of mental states.

[7] Thomas Nagel, "What Is It Like To Be a Bat?" *Philosophical Review* 83 (1974): 435-450 (Cited in Kim 2005, p.94)

§2 The Failures of Alternative Physicalist Accounts

There are, of course, a variety of physicalist responses to both the problem of mental causation and the problem of consciousness. According to Kim, the most successful account—and, perhaps, the *only* successful account—is one whereby the mental is reduced via functional definitions. We will discuss this method of reduction in detail in §3. Before doing so, in order to fully grasp the success of functionally reductive definitions, it will be helpful to examine the failures which Kim takes to undermine alternative physicalist accounts. The scope of this chapter does not allow for an in depth examination of each of these rival positions, but it will be useful to consider the reasons on account of which Kim deems these positions to be failures. With this in mind, we will be in a better position to appreciate the merits that Kim ascribes to a functionally reductive account. *Nonreductive physicalism*, physicalist reduction via *a posteriori identities* and physicalism which invokes *bridge law reductions* all offer responses to the mind-body problem. On Kim's estimation, these responses all fail.

§2.1 Nonreductive Physicalism

Nonreductive physicalism, Kim notes, can be difficult to classify precisely. He writes:

> There is no consensus on exactly how nonreductive physicalism is to be formulated, for the simple reason that there is no consensus either how physicalism is to be formulated or how we should understand reduction. (33)

Despite differences of detail, all (or nearly all) nonreductive physicalists, Kim notes, will affirm the following three positions: the *supervenience* of the mental on the physical, the *irreducibility* of the mental to the physical, and the *causal efficaciousness* of the mental in the physical world. According to Kim, these three theses taken in conjunction with some basic claims of physicalism form an incompatible set, ultimately undermining the tenability of a nonreductive physicalist position. (21-22)

Mind-body supervenience can take a variety of forms, but the supervenience that Kim takes to be central to a nonreductive position is a version of *strong* supervenience. (This is in contrast to *weak* supervenience, which will be discussed later in this chapter.) He defines it as follows:

> *Supervenience*: Mental properties strongly supervene on physical/biological properties. That is, if any system *s* instantiates a mental property M at *t*, there necessarily exists a physical property P such that *s* instantiates P at *t*, and necessarily anything instantiating P at any time instantiates M at that time. (33)

What is crucial about strong supervenience is the claim that the relationship between M and P holds *necessarily*; not only is it the case that instances of P must always be instances of M in the actual world, but in *any* possible world containing instances of P, those instances will be instances of M as well.[8]

Furthermore, on Kim's interpretation of strong supervenience, this covariance should be understood as an indication of the ontological dependence of M on P. He writes,

> I take supervenience as an ontological thesis involving the idea of dependence…*Supervenience*, therefore, is not a mere claim of covariation

[8] This is not to say that M is perfectly correlated with one *and only one* P. Strong supervenience requires that any instantiation of M be an instantiation of *some* P, and that any instantiation of that P must (necessarily) be an instantiation of M as well. If a mental property has multiple supervenience bases, then there will be no single P that is coextensive with M.

> between mental and physical properties; it includes a claim of existential dependence of the mental on the physical. (34)

While one could coherently affirm supervenience without affirming this stronger dependence claim—if, for example, one held the covariation to be the result of pre-established harmony, or of the dependence of the physical on the mental—Kim notes that a *physicalist* should be willing to adopt this weightier position. Most nonreductive physicalists, then, are committed to the claim that all mental properties correlate with and depend upon some physical property or other.

At the same time, nonreductive physicalists of course reject the possibility of *reducing* mental properties to physical properties. They hold instead that, for any mental property M, there is no physical property P such that P is identical to M. They further deny that any physical property P is *coextensive* with some mental property M.[9] Indeed it is the apparent presence of multiple physical supervenience bases for any given mental property that drives many physicalists to adopt a nonreductive approach.[10] Finally, nonreductive physicalists affirm the causal efficacy of the mental in a physical world. That is, they are committed to the reality of mental causation.

Despite this commitment, Kim ultimately concludes that a nonreductive physicalist account of the mind cannot accommodate mental causation. Over the years, Kim has offered a series of "causal exclusion" arguments in favor of this conclusion,

---

[9] Depending on how properties are individuated, these two claims can collapse into one, but they need not. There are plenty of reasons for thinking that two distinct properties might nevertheless be coextensive. One need not accept, for example, that all impossible properties are in fact the same *one* impossible property. Being a round square seems to be a different property from that of being an unmarried bachelor. Likewise, contingently coextensive properties—such as the property of being human and that of being a featherless biped—might not have been coextensive. They, too, seem to be coextensive yet distinct properties.

[10] For an early statement of the Multiple Realization Argument, see: Hilary Putnam, "Psychological Predicates." in W.H. Capitan and D.D. Merrill (eds.), *Art, Mind, and Religion.* (Pittsburgh: University of Pittsburgh Press, 1967)

culminating in the "Supervenience Argument" found in *Physicalism, or Something Near Enough*.[11] This argument will be discussed in detail in Chapter Three. For now, it will suffice to note the broad outline of the argument, which centers upon the following two metaphysical theses:

> *Exclusion*. No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination. (42)
>
> *Closure*. If a physical event has a cause that occurs at *t*, it has a physical cause that occurs at *t*. (43)

The latter thesis, the causal closure of the physical, is a central tenet of physicalism. The precise statement may vary, but most—if not all—physicalists affirm something along the lines of Kim's closure principle. The former thesis, the exclusion principle, is a claim about the nature of causation.[12] Here the claim is that, apart from the occasional coincidence in which two wholly independent and unrelated causal chains manage to simultaneously bring about a single effect, any given event must have one, and only one, sufficient cause at any given time *t*. This principle, though widely affirmed, is not universally accepted among physicalists, particularly those of a nonreductive persuasion. We will examine some of the ways that it has been rejected in Chapter Three. For now, we will bracket these concerns and grant both *Closure* and *Exclusion*.

We are, at last, in a position to see the tension at the heart of the nonreductive physicalist's account. The nonreductive physicalist asks us to suppose that mental states can be causes of physical events. If this is the case, then for any instance of mental

---

[11] Earlier versions of Kim's argument can be found, for example, in his "Making Sense of Emergence," *Philosophical Studies*, 95 (1999): 3-36 and *Mind in the Physical World*. (Cambridge: The MIT Press, 1998): 37-38, 64-67.

[12] It is not at all clear to me that all, or even most, nonreductive physicalists would affirm Kim's exclusion principle. We will examine some of the ways that this principle can be, and has been, rejected in Chapter Three.

causation, *Closure* tells us that there must be a simultaneous instance of purely physical causation culminating in the event that was purportedly also caused by a mental state. *Irreducibility* tells us that the mental cause and the physical cause cannot be identical. In light of *Exclusion*, then, it follows that this must be a case of genuine causal overdetermination. Because we began only with the assumption that *some* instance of mental causation was possible, it follows that *any* instance of mental causation must be a case of genuine overdetermination. This, claims Kim, is unacceptable. If a mental state is to count as a cause, it must make some causal difference. Yet, in order to cause some physical event P, a mental cause M must "somehow ride piggyback on physical causal chains." (48) A world in which physical causes are systematically overdetermined by supervening mental causes is not, claims Kim, a world in which mental causation occurs.

Nonreductive physicalism, then, fails to offer an adequate response to the mind-body problem. Even if it could account for the hard problem of consciousness, such a position would not allow for mental causation; both problems are crucial to the mind-body problem for contemporary physicalism.

§2.2 Identity Theories

In stark contrast to nonreductive physicalism, some physicalists claim that mental properties are not only reducible to, but are in fact *identical with* physical properties. According to the identity theorist, apparent differences between mental and physical properties are just that—they are *apparent*. In recent years, identity theorists (or *type*

physicalists) have looked to Kripkean *a posteriori* identities as the primary means of explicating this position.[13]

In *Naming and Necessity*, Saul Kripke famously argued for the existence of truths that are both *necessary* and knowable only *a posteriori*. To make use of his classic example, he offers the claim "Water is H20" as an instance of an *a posteriori*, necessary truth that we once took to be contingent. Kripke's account relies upon his theory of *naming*, whereby the referent of a concept is determined by ostension and subsequently maintained via a causal chain of speakers. In the case of "water," the referent was determined long before we knew anything about Hydrogen and Oxygen; our concept of water made no mention of, or allusion to, H20. We could never, therefore, have had *a priori* knowledge of the truth that water is H20. It is for this reason that we took this truth to be a contingent one; although we eventually discovered that H20 is the chemical makeup of the stuff we call water, it *seems* as if things might have been otherwise. Nevertheless, given that it is *in fact* H20 that we have been referring to, and that it is—in the actual world—*always* H20 that is picked out by "water," Kripke maintains that the truth is a necessary one. Water—the stuff that is *actually* the referent of "water"—is H20, and cannot but have been H20.[14]

We are now in a position to see how it is that identity theorists make use of Kripke's a posteriori identity statements: Just as we once believed (wrongly) that water

[13] See, for example: Ned Block and Robert Stalnaker, "Conceptual Analysis, Dualism, and the Explanatory Gap." *Philosophical Review*, 108 (1), (1999): 1-46. See also Christopher S. Hill and Brian P. McLaughlin, "There Are Fewer Things in Reality Than Are Dreamt of in Chalmers's Philosophy." *Philosophy and Phenomenological Research,* 59(2), (1999): 445-454.

[14] Saul Kripke, *Naming and Necessity.* (Cambridge: Harvard University Press, 1980) For a more recent way of understanding a posteriori identities, see David Chalmers' account of the "two-dimensional framework." (*The Conscious Mind* (Oxford: Oxford University Press, 1996) pp. 156-165. See also the references listed in fn12.

was only contingently composed of H20, we now believe (wrongly) that pain is only contingently realized by the firing of C-fibers. Just as Kripke showed us the error of our ways with respect to the first belief, additional *a posteriori* identities can likewise liberate us from our current source of confusion. True, our *concept* of pain seems not to have anything to do with C-fibers, or with any kind of neural activity at all. Nevertheless, the *referent* of pain is something physical, and pain is, therefore, *identical to* a physical state. The physical world is all there is—both with respect to substances and with respect to properties and states. This is the identity theorist's claim.

In response to the mind-body problem, then, the identity theorist essentially denies that there *is* a problem. How can there be consciousness in a physical world? Conscious states are physical states, comprised by physical objects and properties. How can the mind exert causal influence in the causally closed physical world? Mental states are physical states; accordingly any account of physical causation will suffice as an account of mental causation as well. The trouble, according to the identity theorist, is that we have mistakenly believed there to be two kinds of things where, in reality, there is only one.

In many ways, it is this rejection of the mind-body problem *as* a problem that lies at the heart of Kim's criticism of identity theories. First, he notes, any such account will fail to close either the explanatory or the predictive gap between the mind and the body. With respect to the explanatory gap, *a posteriori* identities can only tell us that there *never was a gap* to begin with. Suppose we accept the following *a posteriori* identity: (K) Consciousness = pyramidal cell activity. Then, Kim writes:

> Given (K), it no longer makes sense to ask how consciousness emerges out of pyramidal cell activity, not from another sort of neural process, or why there is consciousness just when and where these pyramidal cell activities occur. One is the same as the other, and there is nothing here to explain. (117)

Identities don't *explain* questions of correlations, they *eliminate* them.

To borrow a phrase from David Chalmers, in positing *a posteriori* identities between phenomenal mental states and physical ones, the identity theorist fails to "take consciousness seriously." In *The Conscious Mind*, Chalmers suggests that "To take consciousness seriously is to accept just this: that there is something interesting that needs explaining, over and above the performance of various functions."[15] This is precisely what the identity theorist denies. When asked why a certain experience correlates with some physical state, she claims that the experience just *is* a physical state, and that there is nothing more to be explained. The trouble, notes Kim, is that questions of the sort that Chalmers raises seem to be "perfectly intelligible." (119) For this reason, anyone inclined toward taking consciousness seriously will simply reject the identities posited by the identity theorist. Absent some independent evidence in favor of the *truth* of these identities, there mere *usefulness* will not suffice as an adequate answer to the mind-body problem.[16]

Even if the identity theorist could persuasively argue that the explanatory gap is unproblematic, the *predictive* gap would remain unaddressed. Kripkean identities are *a posteriori*. By definition, we cannot know the truth of some Kripkean identity without first having observed the relevant correlation. As such, we could never be in a position to posit a Kripkean identity between, say, the neural structure of a bat and some phenomenal experience. We might attempt to *infer* something about the experience of the bat by appealing to our own experiences, but we do not have the requisite epistemic access

[15] (Chalmers 1996, p. 168)

[16] For a more detailed treatment of Kim's response to the identity theorist, I refer the reader to Chapter 5 of (Kim 2005). There, Kim examines and rejects arguments based upon "inference to the best explanation" and explanatory success as insufficiently motivated.

which we would need in order to posit a genuine Kripkean identity. Consider once again the question posed by the emergentists: Given complete physiological knowledge of a creature, can we predict anything about its phenomenal conscious states? To this question, the identity theorist must respond that no, we cannot. One crucial question then remains: If reductive physicalism is true—if all of the facts about the world are reducible to physical facts—then why is knowledge of the physical facts insufficient for knowledge of all of the facts?[17]

Identity theory, then, cannot respond to the problem of consciousness. Despite the difficulties that result from an attempt to *identify* all mental states with physical ones, Kim concludes that reduction of *some* kind is ultimately necessary if the physicalist is to save mental causation. Nonreductive accounts, for reasons we have discussed, must inevitably succumb to systematic overdetermination or the causal impotence of mental states. In light of this, Kim writes "If we want robust mental causation, we had better be prepared to take reductionism seriously." (22) That said, reduction can take a variety of forms. One method of reduction, a method that initially drew a great many followers, is reduction of the mental to the physical via "bridge laws."

[17] For the canonical statement of the so-called Knowledge Argument, see: Frank Jackson's "Epiphenomenal Qualia," *Philosophical Quarterly* 32 (1982): 127-136.

§2.3 Reduction via Bridge Laws

Bridge law reduction was first suggested by Ernest Nagel in *The Structure of Science*.[18] On this method, entire theories were to be reduced to more fundamental ones through the "bridging" of the individual predicates of the higher-level theory to predicates in the base theory. Kim writes,

> Nagel reduction requires that each primitive predicate, M, of the theory being reduced be connected with a predicate, P, of the base theory in a 'bridge law' (or 'bridge principle') of the form:
>
> For any $x_1 \ldots, x_n$, $M(x_1, \ldots x_n)$ if and only if $P(x_1, \ldots x_n)$. (98)

For example, let M be the predicate "57 degrees Celsius" and P the predicate "motion of speed K." For there to be a true bridge law connecting the two, it must be the case that all objects are 57 degrees Celsius when, and *only* when, their molecular motion is of speed K. Given this bridge law, M, a predicate in terms of temperature, can be reduced to P, a predicate in terms of molecular motion.

When all of the primitive predicates of a theoretical language are bridged in this way to predicates of a more fundamental theory, the higher-level theory can itself be reduced to the more fundamental one. Using the language and laws of the reduction base coupled with the bridge laws, any truth of the higher-level theory—including the *laws* of that theory—can now be expressed wholly in the language of the base theory. Suppose, for example, that all of the predicates of sociology were reducible via bridge laws to predicates of biology. Then any law of sociology could be stated purely in biological terms. In this way, sociology itself could be reduced to biology; there would be nothing *to* sociology over and above the truths of biology.

[18] (New York: Harcourt, Brace and World, 1961), Ch. 11. Cited in (Kim 2005, p98.)

If available, bridge law reductions could offer a response to the problem of mental causation and, to a lesser extent, to the explanatory component of the problem of consciousness. With respect to the former, suppose that all mental predicates were reducible (via bridge laws) to predicates in the language of a completed physics. A solution to the problem of mental causation would, in this case, be rather straightforward. Every instance of mental causation could be restated wholly in the language of physics; mental causation would be but a subspecies of physical causation. Given this scenario, there would be no problem of mental causation. With respect to the problem of consciousness, however, things get a bit more difficult on this account. Consider the question, "Why does pain arise whenever C-fibers are firing?"[19] In order to answer this question using bridge laws, we must be able to posit a bridge law along the following lines: "For all subjects S, S is in pain if and only if S's C-fibers are firing." There is a sense in which such a bridge law could tell us why pain arises whenever C-fibers fire—simply stated, *because* there is a true bridge law correlating the two.

Yet there is another, strong sense in which this answer fails to be at all explanatory. Bridge laws are supposed to be contingent, and they are knowable only through empirical means. As such, unlike Kripkean identities, there is nothing incoherent about asking *why* any given bridge law is true. Indeed, in asking why pain arises whenever C-fibers fire, it seems that this is precisely what we *are* asking. Kim writes,

> In using the bridge laws as auxiliary premises of reductive derivations, the Nagelian reductionist is simply assuming exactly what needs to be derived and explained if we are to answer the explanatory questions raised by Huxley, James and the emergentists—that is, if we are to close the explanatory gap or solve the hard problem of consciousness. (100)

[19] As Kim notes, the pain/C-fiber correlation is the oft-cited product of "philosophers' fictional neurophysiology." (Kim 2005, p.13)

Consider a proposed reductive explanation of a particular occurrence of pain. Suppose we want to know why a pinprick to the finger causes the experience of pain. An answer in terms of bridge law reductions will run along the following lines: The event which we describe as "a pin prick to the finger" can (in principle) be restated entirely in the language of a completed physics. Likewise, the event which we call "experiencing pain" can be similarly restated. In both cases, we need only appeal to bridge laws linking predicates such as "being pricked by a pin" and "feeling pain" to fundamental physical predicates such as, "being in physical state S" and "undergoing the firing of C-fibers." These bridge laws, coupled with a purely physical account of why physical state S causes C-fibers to fire, will suffice as an explanation of why a pin-prick to the finger causes the experience of pain.

What will *not* be explained, however, is the truth of these bridge laws themselves—most notably, the bridge law correlating the phenomenal experience of pain with the firing of C-fibers. Instead, in order to get any explanation of an occurrence of pain, the truth of this bridge law will have to be *assumed*. Bridge law reduction, then, cannot tell us why any particular conscious state correlates with some physical state; at best, it can tell us why this conscious state occurs when it does *given that* it is so correlated. This is an answer to some question, to be sure, but not to the question posed by the problem of consciousness.

Additionally, as Kim notes, reductive explanations in terms of bridge laws are not actually *reductive* in the right way. After all, any such explanation will have to appeal to bridge laws, and bridge laws cannot be stated entirely in the language of the reduction base. Instead, insofar as they are *bridge* laws, they must make reference to the predicate *to be reduced*. Thus, bridge law reductions appeal to a base theory that includes

predicates of the higher-level theory, a theory comprised of the original reduction base *supplemented with* the bridge laws. For this reason, bridge law reductions must rely upon the expansion of the base domain, and so are not fully reductive, and must assume the truth of the bridge laws, and so are not explanatory. (99-100)

Finally, because bridge laws are only knowable through empirical investigation, bridge law reduction cannot address the *predictive* component of the problem of consciousness. Like Kripkean identities, bridge laws can only account for correlations between conscious states and physical states once the conscious states have already been observed. Simply stated, no amount of purely physical knowledge could ever generate physical-to-mental bridge laws. For this reason, bridge law reduction will be utterly useless with respect to the predictive problem of consciousness.

In attempting to respond to the mind-body problem(s) for contemporary physicalism, nonreductive physicalism, physicalism in terms of *a posteriori* identities, and reductive physicalism via bridge laws all, according to Kim, fail to address the problem in one form or another. Nonreductive physicalism cannot meet the demands of the problem of mental causation; instead, it collapses into epiphenomenalism or appeals to the systematic overdetermination of mental causes. Kripkean identities can say nothing about the explanatory gap, apart from claiming that there *is* no such gap, and are incapable of addressing the predictive problem of consciousness. Bridge law reductions are equally impotent in the face of the predictive problem of consciousness, and—like Kripkean identities—assume the truth of the correlations for which the explanatory problem of consciousness seeks an explanation. If the physicalist is to address the problem of mental causation and the problem of consciousness, she must look elsewhere for a solution. In what follows, we will examine the ability of functional-reductions to do

this work. According to Kim, functional reductions can provide an answer to the problem of mental causation and to the problem of consciousness (for the most part) in a way that none of these alternatives could.

§3.1 Reduction via Functional Definitions

Kim's functional reduction is a three-step process.[20] In order to functionally reduce a property, we must first construct a definition of the property in terms of its causal role. Kim uses the example of "being a gene" as a property that can be reduced in this manner. (101) "Being a gene" can be defined as "being a mechanism that encodes and transmits genetic information." (101) With this definition in hand, we then seek to find the properties or mechanisms in the reduction base that play this causal role. Whatever it is that encodes and transmits genetic information comprises a *realizer* of "being a gene." Finally, we construct an account that explains how it is that these realizers play the causal role that they do. In explaining how the realizers of "being a gene" encode and transmit genetic information, we will thereby have constructed an account of how it is that *genes* encode and transmit genetic information.

According to Kim, a physicalist who wishes to respond to the problem of consciousness and the problem of mental causation—that is, who is concerned with the mind-body problem—ought to invoke functional reductions of the mental to the physical. Consider first the problem of mental causation. If mental properties can be given functional definitions in terms of the physical domain, then mental causation can be

[20] As Kim notes, David Chalmers and Joseph Levine both advocate similar methods of functional reduction. See: Joseph Levine, *Purple Haze* (Oxford: Oxford University Press, 2001) and Chalmers, 1996.

reductively explained. Take, for example, the mental property "being amused." Suppose I want to know how it is that my being amused is causally responsible for my laughing. If we define "being amused" as "having some physical property or other that is apt to trigger smiles or laughter," then we can construct an explanation in terms of this definition; we need only find the realizers of this causal role and explain how it is that they play the role that they do. In doing so, we will have explained how it is that "being amused" causes laughter. In short, functional definitions allow us to explain the causal relevance of a mental property by appealing to the causal activity of the physical realizers of the functionally reduced mental property.[21]

If a mental property is functionally reducible, and if the realizers of that property are causally relevant, then the mental property can itself be deemed causally relevant in virtue of those realizers. Furthermore, according to Kim, only the first step of the three-step process need be completed before we can be justified in ascribing reducibility to a property. "That a property is functionalizable—that is, that it can be defined in terms of a causal role—is necessary and sufficient for functional reducibility." (165) If we wish to claim that we have *reduced* a property, then we had better find the realizers. However, for *reducibility*, we need only construct a functional definition. In order to ascribe causal

[21] There are some issues here worth considering more carefully. For example, what do we say about the reduced property once the reduction has taken place? Does "being amused" remain a property in its own right, or is it replaced by the functionally defined, lower level property? On Kim's account, it seems we must conclude either that (a) the reduced mental properties are just a subspecies of physical properties, and functional definitions somehow allow us to *see* the (already physical) property as it really is, or (b) there are, strictly speaking, no mental properties, but only mental concepts which we took to refer to mental properties but which, ultimately, refer to physical ones. I will discuss this in greater detail later in this chapter. For now it will suffice to note that, according to Kim, the causal activity of the physical realizers of a functionally defined mental property can be invoked to explain the causal relevance of the reduced mental property.

relevance to a mental property, then, we need only be able to construct a functional definition of that property in terms of the physical domain.[22]

Fortunately, Kim notes, we have good reason to believe that intentional and cognitive mental states can be given functional definitions. In support of this claim, Kim offers the following illustration:

> Consider a population of creatures, or systems, that are functionally and behaviorally indistinguishable from us, and, in general, observationally indistinguishable from us…In particular, they exhibit similar linguistic behavior; as far as we can tell, they use language as we do for expressive and communicative purposes. If all this is the case, it would be incoherent to withhold states like belief, desire, knowledge, action, and intention from these creatures. (165)

Belief, desire, knowledge, action and intention are, according to Kim, exhaustible by their physical manifestations. If the physical manifestations of these mental states are present, we can conclusively determine that the mental states are present as well; indeed, that is all that we *mean* when we say that the mental state is present. (As is no doubt clear, Kim is not persuaded of the possibility of a so-called "Zombie world.")[23] Because they are so exhaustible, they can be given definitions in terms of physical realizers. We may not have the realizers in hand at the moment, but we can rest assured that they are in-principle available, and that these mental states are reducible to the physical. In this way, functional reductions can enable the physicalist to offer a plausible response to the problem of mental causation.

---

[22] This, of course, assumes that the realizers of the property will not turn-out to be epiphenomenal physical states. Given the scarcity of epiphenomenal physical states, this seems to be a safe assumption. We could, alternatively, resist ascribing causal relevance to a mental property until after at least some of its (causally active) realizers have been identified.

[23] See, for example, his discussion on p27 and p169.

Furthermore, functional reductions enable the physicalist to respond to both the explanatory and the predictive components of the problem of consciousness.[24] Kim offers the following example of a reductive explanation, supposing that pain has been given a functional definition:

> Why is Jones in pain? Because to be in pain is to be in some state that is apt to be caused by tissue damage and apt for causing winces and groans, and Jones is now in neural state N, which, as it happens, is a state apt to be caused by tissue damage and apt for causing winces and groans. (112)

Functional definitions can thus be employed to explain why a subject is in a given mental state at a given time. Once a mental state has been given a functional definition, and one or more of its realizers have been identified, its occurrences can be explained in terms of the occurrences of its physical realizers.

Additionally, unlike Kripkean identities, functional definitions can explain why a certain mental state correlates with a particular physical state, and can do so without rendering the question of correlation nonsensical. Again, Kim offers an example:

> A system, *x*, is in neural state N at *t*.
> Neural state N satisfies causal role C (in systems like *x*).
> Having pain $=_{df}$ being in some state satisfying causal role C.
> Therefore, *x* is in pain at *t*. (112)

Because pain is not *identified* with neural state N, there is nothing tautological about saying that every instance of N is an instance of "being in pain." The crucial claim of the third premise is that N qualifies, in virtue of the causal role that it plays, as one of the realizers of the concept "being in pain;" it does not follow that neural state N *is* pain. By way of analogy, it is not trivial to note that a particular round, heavy rock is a paperweight, even if the rock is seen sitting on my desk on a stack of papers. In doing so,

[24] More accurately, functional reductions—insofar as they are *available*—can help with both the problem of mental causation and the problem of consciousness.

we are asserting that this rock is a realizer of "paperweight", and is so in virtue of its functional role. For this reason, a functionally reductive explanation of the correlation of a mental state and a neural state need not undermine the validity of asking *why* the correlation obtains, just as a functional definition of "paperweight" need not render senseless the question "Why is *that* rock a paperweight?".

In contrast to bridge law reductions, functionally reductive explanations also avoid the pitfall of appealing to properties outside of the reduction domain; that is, they can be sufficiently *reductive*. In evaluating the failure of bridge-laws, Kim formulates the following constraint upon reductive explanations:

> (R) The explanatory premises of a reductive explanation of a phenomenon involving property F (e.g., an explanation of why F is instantiated on this occasion) must not refer to F. (105)

He then goes on to suggest that (R) might be strengthened, so that a reductive explanation of a phenomenon involving F could refer neither to F nor to any phenomenon at the level of F. (106) To see why functionally reductive explanations do not violate either formulation of constraint (R), Kim asks us to note that "having pain" in the third line of the sample explanation just given does not refer to some mental *property*, but is instead a definition of the *concept* "being in pain," or the term "pain." (111) For this reason, despite initial appearances, a functionally reductive explanation need not appeal to the property that it aims to reduce.[25] It will suffice to mention the term, or concept, that we associate with a higher-level phenomenon, and offer a definition of that term wholly in the language of the reductive base theory.

[25] Again, one might wonder whether or not there *are* mental properties on Kim's account, or if instead there are *only* these mental concepts which refer to physical properties. This will be discussed in §6 of this chapter.

Functionally reductive explanations, then, have an advantage over rival physicalist accounts when it comes to closing the explanatory gap. Where Kripkean identities require the physicalist to deny that there ever really was a gap to begin with, functionally reductive explanations offer a solution to the mystery of correlation without denying that there was a mystery to begin with. Similarly, where bridge-law reductions require the physicalist to expand the reductive base to include certain higher-level properties, functional reductions make no such demands. In this way, the physicalist who invokes the functional method of reduction can offer reductive explanations of mental phenomena that are both genuinely *reductive* and genuinely *explanatory*.

Finally, functional definitions also allow the physicalist to respond to the *predictive* problem of consciousness. Where Kripkean identities are *a posteriori*, functional definitions are the result of conceptual analysis, and so are *a priori*. For this reason, they allow for predictions of unobserved mental states. If pain just means "being in a state apt to be caused by tissue damage and apt for causing winces and groans," then we can reasonably conclude that a bat, or a dog, who has experienced tissue damage will also be in pain. If the animal then proceeds to wince and groan, we will have further confirmation of this fact. Given full knowledge of the physical states of a creature, coupled with functional definitions of all known mental states, we can deduce the mental states of that creature.

It is important to note that such a process would result in *conclusive*, and not merely probabilistic, evidence; if we have the correct definition of a mental state, and have identified a physical realizer of that state, we can be *certain* of the presence of the mental state. By claiming that a functional definition is an adequate one, we are claiming that there is *nothing more to* the mental state than the functional role detailed in the

definition. It is, therefore, nonsensical to suppose that a correctly defined mental state might be absent despite the presence of one of its realizers.

To summarize the benefits of this account which have so far been considered, note that functional reduction manages to avoid many of the difficulties beset by alternative physicalist accounts. Unlike nonreductive physicalism, functional reduction can account for mental causation without collapsing into systematic overdetermination. Unlike identity theory and bridge law reduction, functional reduction can provide the physicalist with a response to both the explanatory and the predictive gap, and can do so without rendering the questions underlying the explanatory gap nonsensical. Finally, functional reduction does not require any expansion of the reduction base; given the physical domain alone, functionalized mental concepts can be wholly accounted for.

§3.2 Functional Definitions: The Qualia Problem

If all of our mental concepts, or properties, could be functionally defined, then the physicalist who invoked such reductions could consider both the problem of mental causation and the problem of consciousness a thing of the past. Unfortunately, as Kim concedes, not all mental concepts can be functionalized. In particular, qualitative mental states, or *qualia*, cannot be given functional definitions. They are not exhaustible by their physical manifestations, and so resist reduction.

To demonstrate the irreducibility of qualia, Kim appeals to the metaphysical possibility of a qualia inversion scenario. The problem of qualia inversion runs roughly as follows: While you and I might exhibit identical behavior when viewing a ripe tomato,

my qualitative experience of the tomato might resemble the experience that you typically have when viewing a tomato that has not yet ripened. Intrinsically, your qualitative experience of red might by identical to my qualitative experience of green. Furthermore, this could be the case for all of our qualitative experiences. If our qualia spectrums were fixed, and so consistent for the duration of our lives, then this difference could not be physically detected.[26] If such inversion is metaphysically possible, then

> Two perceivers who behave identically with respect to input applied to their sensory receptors can have different sensory experiences. If that is true, qualia are not functionally definable; they are not task-oriented properties. (170)

Because the nature of a quale is not exhausted by its functional role—insofar as it *has* a functional role—a quale cannot be adequately captured by a functional definition. Qualia, then, are not candidates for functional reduction. Furthermore, the fact that qualia contain features that go beyond their physical manifestations has significant ramifications for physicalism as an ontological thesis: qualia are irreducible to physical states, and so are not a part of the physical world.

In light of this "mental residue," Kim draws the following conclusion: "Global physicalism is untenable…There is a possible world that is like this world in all respects except for the fact that in our world, qualia are distributed differently." (170) Still, having shown in his Exclusion Argument that reducibility is necessary for causal efficacy, Kim takes comfort in the fact that irreducible qualia must be epiphenomenal. Qualia, he writes, "stay outside the physical domain, but they make no causal difference and we won't miss them." (173) Furthermore, similarities and differences among the qualitative experience of any given individual *may* turn out to be functionalizable, so certain aspects of qualia

[26] It is, of course, crucial to the thought-experiment that the qualia difference not be traceable to any physical difference, as it is in the case of people who are simply color-blind.

can perhaps be saved. Physicalism, he concludes, is "not the whole truth, but it is the truth near enough, and near enough should be good enough." (174)

§4. Kim's Functional Reduction: Tying up the Loose Ends

The work accomplished in *Physicalism, or Something Near Enough* is significant and far-reaching, and Kim's use of functional reductions to respond to the mind-body problem is impressive in its level of detail. However, as is to be expected in a project of such complexity, his own positive account ultimately leaves a number of questions unanswered. In concluding *Physicalism, or Something Near Enough*, Kim writes:

> There are many issues that need to be sorted out in more detail and with greater care and precision; among them are (1) the functional reducibility of cognitive/intentional states, (2) the functionalizability of qualia differences and similarities, (3) whether qualia epiphenomenalism is consistent with the assumed fact that the subject of experiences is cognitively aware of them and is able to make reports about them, and (4) the question whether it is possible to combine qualia epiphenomenalism with full causal efficacy of qualia similarity and differences. (173-174) (*my numbering*)

In the next two sections, I wish to shift my focus to these, and other, outstanding issues. I suggest that, until we know how Kim would have us address the remaining questions, it is difficult to know how we should assess his account as a whole. I will not presently treat (1), as §6 will be devoted, in part, to the question of how exactly we should understand the reduction of those mental states that are eligible for functionalization. Instead, I will begin with a discussion of the questions raised in (2)-(4).

Before addressing questions (2) and (4), two clearly interrelated concerns, I would like briefly to consider question (3): can we make sense of a subject's being aware of

epiphenomenal qualia? How might a person *become* aware of them? This much, at least, is clear: If we can become aware of epiphenomenal qualia, it must be through noncausal means. Epiphenomenal qualia can exert no causal influence over us, and so *they* can of course not cause us to know what it's like to experience a given quale. They likewise can exert no *indirect* causal influence over us, for they cannot cause anything else to serve as a causal intermediary. If Kim wishes to defend the claim that the subjects of experience can become cognitively aware of qualia, he must do so by invoking some noncausal account.

This is a fairly obvious point given the epiphenomenal nature of qualia, but I suggest that it is nevertheless worth noting. Adopting a noncausal account of knowledge acquisition, perhaps something along the line of William Alston's "appearing" relation, is not, in and of itself, problematic.[27] Perhaps qualia somehow "appear" to us, or perhaps we are directly acquainted with qualitative properties; neither is an obviously troubling position. That said, Kim has traditionally been a staunch advocate of the centrality of causation to knowledge and explanation.[28] One might be able to affirm noncausal qualia *awareness* without affirming noncausal knowledge, but that seems a difficult road to take.[29]

---

[27] William Alston "Perception and Representation" *Philosophy and Phenomenological Research* Vol. LXX, No. 2, March 2005 p.257 Leopold Stubenberg alerted me to this account.

[28] See, for example, "Mechanism, Purpose, and Explanatory Exclusion," *Philosophical Perspectives* 3 (1989): 77-108.

[29] In addition to his past philosophical commitments, I suggest that there are more pressing reasons for which Kim might be inclined to reject a noncausal account of qualia knowledge and, indeed, even of qualia awareness. In Chapter Three of *Physicalism, or Something Near Enough,* Kim raises the Pairing Problem against substance dualism. I will not treat this problem in any detail here. (I discuss the problem in detail in Chapter Two of my dissertation, "Physicalism, Dualism and the Mind-Body Problem.") Succinctly stated, Kim argues that an immaterial mind cannot enter into causal relations because it cannot be uniquely paired with any given cause or effect. More importantly for our purposes, he explicitly considers the possibility of appealing to intentional psychological relations in order to pair minds with objects. There, he draws the

That said, there is nothing *prohibiting* Kim, or any other advocate of functional reduction, from embracing a noncausal account of qualia awareness. (In fact, David Chalmers affirms a position very much like Kim's and explicitly notes that our awareness of qualia *must* come through noncausal means.)[30] For the purposes of this discussion, then, I will assume that Kim can successfully address question (2). In light of this concession, what follows for the functional reduction of qualia differences? Will the mere awareness, or knowledge, of qualitative properties somehow enable us to functionally reduce difference or similarities among them?[31]

Perhaps, but I can't see how. To be honest, I can't say that I really understand what it would *mean* to reduce—functionally or otherwise—a difference or a similarity. In

---

following conclusion: "To pick out some concrete thing outside us, we must be in a certain cognitive relation to it; we must perceive it somehow and be able to single it out from other things near and around it…Ultimately, these intentional relations must be explained on the basis of causal relations." (81)

It seems, then, that Kim is generally opposed to noncausal relations of the sort necessary for qualia awareness; instead, he suggests that perception, and other intentional relations, implicitly rely upon an established causal relation. If Kim nevertheless chooses to affirm the claim that nonphysical, epiphenomenal qualia can somehow be made known to us in a noncausal way, he will inevitably open the door to a variety of dualistic responses to the pairing problem.

[30]In *The Conscious Mind*, he writes, "A property dualist should argue that a causal theory of knowledge is not appropriate for our knowledge of consciousness, and that the justification of our judgments about consciousness does not lie with the mechanisms by which those judgments are formed." (193) And "If one takes consciousness seriously, then one has good reason to believe that a causal or reliabilist account of our phenomenal knowledge is inappropriate." (Chalmers 2005, p.196)

[31] In what follows, I will assume that Kim's irreducible qualia are best understood as irreducible mental properties. When introducing the arguments found in *Physicalism, or Something Near Enough,* Kim writes that "phenomenal mental properties are not functionally definable and hence functionally irreducible." (29) There, he clearly takes irreducible qualia to be mental properties. Similarly, in the conclusion of this work, he explicitly refers to qualia as "qualitative properties of consciousness." (174) It seems clear, then, that qualia are to be understood as properties. It is worth noting, however, that if qualia are properties, then property dualism is true; this despite Kim's rejection of property dualism in the conclusion of *Physicalism, or Something Near Enough*. (See, for example, pp158-159) A careful reading suggests that what Kim really rejects is not dualism *per* se, but rather the ability of a more traditional property dualism to account for mental causation. (Because the nonphysical properties on Kim's account are supposed to be epiphenomenal, the dualism of properties posited by Kim is not intended to have any significant bearing on the problem of mental causation.) If that is the case, however, and I am right to conclude that qualia are properties, then it seems worth noting that Kim's near physicalism is, in fact, a property dualism.

defending the potential possibility of a reduction of this sort, Kim offers the following analogy:

> Consider traffic lights; everywhere in the world, red means stop, green means go, and yellow means slow down. But that is only a convention, the result of a social arrangement; we could have adopted a system according to which red means go, green means slow down, and yellow means stop, or any of the remaining combinations. That would have made no difference to traffic management. What matters are the differences and similarities among colors, not their intrinsic qualities. (172)

This much is certainly true: there is nothing intrinsically stop-inducing about a red qualitative experience, nor is a yellow experience the sort of thing that just tends to make one slow down. Surely any color combination would serve this purpose equally well, and it is the fact that there are three distinctly colored lights that accounts for whatever causal role traffic lights play. In this sense, there seems to be some causal function performed by the combination of lights that could not be played by any single light in the absence of others.[32]

There are, however, two points that need to be made in response to this analogy. First, it is not yet clear that this color-difference can be functionalized in a way that leads to a single causally efficacious, functional property. For that to be the case, we must be able to carefully articulate some specific causal role that the color-difference realizes.[33] Perhaps that can be done, but it has not yet been done. If the analogy is to help bring qualia differences into the causal realm, then this must be possible.

---

[32] I do not mean to suggest that all three colors must be *present* in order for this function to be fulfilled; surely a blinking red light has its desired effect, despite the fact that it is not physically accompanied by a red and a green light. Still, it is only because we have also seen green and yellow lights, and been taught the role that each plays, that we ascribe the meaning that we do to red lights.

[33] Additionally, the resulting functional property will still need to contend with the problems raised by the causal exclusion argument in the preceding discussion.

Secondly, and more importantly, if such a reduction can be obtained, it need not follow that the *phenomenal experiences* that accompany the perception of red, yellow and green lights have anything to do with the matter, causally speaking. Indeed, if we begin with the assumption that qualia are epiphenomenal, we ought to withhold any attribution of causality to relationships among qualia until we have some reason for believing them to be functionalizable, or otherwise reducible. If, instead, we assume that the causally relevant functional property in the above scenario involves differences among *qualia*—rather than, say, differences among whatever *physical* states accompany our color perception—we assume that qualia differences can be functionalized.

There is, as well, a positive reason to believe that the resulting functional property would not capture differences among qualia themselves: when coupled with the possibility of a qualia inversion scenario, this proposal leads to great difficulties. Suppose that there is a world, W*, that is just like this world, only in W* my qualia spectrum is inverted with respect to my spectrum in the actual world.[34] While and Dolores* and I lead lives that are externally indiscernible, Dolores* has a green color experience whenever I have a red one. If the proposed functional property *does* functionalize differences between color experiences, rather than between physical perceptual states, then one of two things must be the case. Either (a) the property responsible for my stopping at a red light is distinct from that responsible for Dolores*'s stopping, or (b) the property is the same, but Dolores* stops at green lights instead of red ones.

The latter option is, of course, not really an option; we are assuming that there is no observable difference between Dolores* and myself. If the properties were the same,

[34] If we do not wish to posit two physically indiscernible possible worlds, we can instead add some minor difference in the causal history of the world prior to my existence.

however, then the property responsible for causing a person to stop at a red light would, in both worlds, be a functionalization of the difference between the red experience and the yellow and green experiences. Dolores*, of course, has the red experience when a green light (or what *I* would call a green light) is illuminated, so this difference would be realized for her in situations in which I see a green light. The red experience, or the "red-difference" experience, cannot play the causal role in W2 that it does in the actual world if our behavior is to be indiscernible. It must therefore be the case that the properties are not the same after all.

Yet this is an odd move to have to make. After all, the functionalized property makes no reference to phenomenal experience. It is expressible in wholly physical language, and Dolores* and I do not differ in any physical respects. If we assume, nevertheless, that there are *two* properties here, F1 and F2, then note what follows: F1 and F2 share precisely the same functional definition, and are realized *in both worlds* by precisely the same physical states, yet they are distinct properties. So stated, this distinction seems unmotivated, if not ad hoc.

This is surely a conclusion to be avoided. Furthermore, it can *be* avoided; one need only assume that the functional property responsible for traffic light behavior—insofar as there is some such property—is a functionalization of the differences between three physical perceptual states, rather than phenomenal ones.[35] Given the epiphenomenal nature of qualia, this is what we ought to have expected all along. Just as it is not really the *experience* of pain that causes me to wince, it's not really the *experience* of red that

[35] It will not do to appeal to the fact that phenomenal experience seems to be causally relevant. After all, my experience of pain sure seems to cause me to wince, but if phenomenal pain is epiphenomenal than we know this not to be the case. Seeming to be causally efficacious, then, cannot suffice as evidence of causal efficaciousness.

causes me to stop. In light of this, it seems unlikely—if not impossible—that the *difference* between my red experience and my green experience could be any more causally relevant than the experiences themselves. It seems clear, then, that even if the physicalist could somehow functionalize similarities or differences of *some* variety, it does not follow that *qualia* similarities or differences are functionalizable.

§5. The Question of Qualia Supervenience

Before revisiting the problems of mental causation and of consciousness, I wish to consider one additional outstanding question not explicitly mentioned by Kim. I suggest that how Kim responds to this question will significantly impact the overall character of his ontology, and will therefore be of great importance when evaluating his position as a whole. The question is this: Assuming that qualia are nonphysical properties, what relation do they bear to the physical world? More specifically, do they, or do they not, supervene on physical properties?

Recall Kim's definition of mind-body supervenience as formulated in his discussion of nonreductive physicalism:

> *Supervenience*: Mental properties strongly supervene on physical/biological properties. That is, if any system *s* instantiates a mental property M at *t*, there necessarily exists a physical property P such that *s* instantiates P at *t*, and necessarily anything instantiating P at any time instantiates M at that time. (33)

If this is what is meant by "supervenience", then qualia do not supervene on physical properties. If they did, then the qualia inversion scenario would be metaphysically

impossible, and qualia would be functionally reducible just like any other mental property.

To see why, suppose that John and Jim have qualia spectrums that are inverted with respect to one another. John is an inhabitant of the actual world. Jim might be as well, though he could also be an inhabitant of a distinct possible world; it will not matter for the purposes of this illustration. On this assumption, John and Jim could be in precisely the same physical state and yet be experiencing different qualitative ones. Suppose that, like Dolores* and myself, John and Jim are red-green inverted such that in whatever circumstances John experiences red, Jim experiences green.

If qualia strongly supervened on physical properties, then this could not be the case. On the supervenience hypothesis, every time John instantiates "experiencing red," this instantiation will necessarily be accompanied by some physical property P (or other) such that *any* instantiation of P is *of necessity* an instantiation of "experiencing red." But then any instantiation of P *by Jim* must be an instantiation of "seeing red;" that is, Jim must see red in any physical scenario in which John would see red. If qualia strongly supervene on physical properties, then the qualia inversion scenario is impossible. Because Kim takes the qualia inversion scenario to *be* metaphysically possible, he must reject the strong supervenience of qualia on physical properties.

This is not to say that Kim must reject *all* forms of qualia supervenience. He might instead affirm the *weak* supervenience of qualia on the physical world. In "'Strong' and 'Global' Supervenience Revisited," Kim defines weak supervenience as follows: Where A and B are two sets of properties, A *weakly supervenes* on B just in case "Necessarily, for any x and y, if x and y share all properties in B, then x and y share all

properties in A—that is, indiscernibility in B entails indiscernibility in A."[36] If qualia weakly supervene on physical properties, then physically indiscernible subjects *within a given possible world* must also be qualitatively indiscernible. This is the crucial difference between weak and strong supervenience: where the latter requires covariation across *all* possible worlds, weak supervenience limits this requirement to individual worlds. For this reason, the weak supervenience of qualitative states on physical ones does not preclude the metaphysical possibility of spectrum inversion.[37]

Let us assume, then, that Kim would affirm the weak supervenience of qualitative mental states on physical ones.[38] If that is the case—if, that is, Kim wishes to affirm weak qualia supervenience while denying strong qualia supervenience—then Kim will have to contend with a challenge raised by Simon Blackburn in "Supervenience Revisited."[39] There, Blackburn examines the tenability of any position that both affirms the weak supervenience of one set of properties on another while denying the strong supervenience of those properties. He notes that such a position requires a "ban on mixed worlds," and that a ban of this sort must be rationally justified in some way.[40]

---

[36] in *Philosophy and Phenomenological Research* Vol. 48, No.2, (Dec 1987) p.315

[37] Weak supervenience is *not* consistent with the possibility of two physically indiscernible subjects differing with respect to qualia *in a given world*, but there is no indication that Kim affirms this possibility.

[38] Indeed, Kim himself has linked the acceptance of the possibility of qualia inversion with weak qualia supervenience. In "Concepts of Supervenience," he writes that weak supervenience "may be held by those who take the attribution of mental states as just another case of positing theoretical explanatory states (relative to, say, behavior), and who take the possibility of the 'inverted spectrum' seriously." "Concepts of Supervenience," *Philosophy and Phenomenological Research* 45 (1984): 153-176. Reprinted with permission in Kim, Jaegwon *Supervenience and Mind.* (Cambridge: Cambridge University Press, 1993) p.63 (Page numbers refer to Kim, 1993)

[39] In Hacking, Ian (ed.) *Exercises in Analysis.* (Cambridge: Cambridge University Press, 1985) pp.47-67

[40] (Blackburn 1985, p.53)

To see what Blackburn has in mind, suppose that qualia only weakly supervene on physical properties. Let Q be a qualitative property and P1 one of its supervenience bases in world W1. Any instantiation of P1 in W1, then, will be an instantiation of Q. Because strong supervenience fails, however, there must be a world (W2) in which P1 is *not* a supervenience base of Q, though both P1 and Q are instantiated.[41] In W2, there will be some other physical property (P2) that is a supervenience base of Q, but instantiations of P1 will not be instantiations of Q.

If qualia only weakly supervene on the physical, then, there will be possible worlds in which instantiations of P1 *always* give rise to instantiations of Q, and there will be worlds in which instantiations of P1 *never* give rise to instantiations of Q, but there will be *no* worlds in which instantiations of P1 *only sometimes* give rise to instantiations of Q. If there were "mixed worlds" of this variety, then even the weak supervenience of qualitative properties on physical ones would fail. The question, according to Blackburn, is *why* we should think that there are no worlds of the third kind. He asks:

> Why should the possible worlds partition into only the two kinds, and not into the three kinds? It seems on the face of it to offend against a principle of plenitude with respect to possibilities, namely that we should allow any which we are not constrained to disallow.[42]

If an advocate of merely weak supervenience were to offer some positive account whereby "mixed worlds" were shown to be impossible, then of course the so-called "ban on mixed worlds" could be a justified one. Absent some such account, however, it's difficult to see why we should accept the stipulation that no such worlds exist.

---

[41] Strictly speaking, there must only be *some* qualitative property for which the supervenience base differs across possible worlds. Assume for the example that Q is one of the properties that displays the failure of strong supervenience.

[42] (Blackburn 1985, p.53)

Blackburn's challenge can be further strengthened by Kim's own stated position on supervenience with respect to the mind-body problem. In "Postscripts on Supervenience," Kim makes the following claim:

> Any physicalist who believes in the reality of the mental must at a minimum accept pervasive psycho-physical property covariance (in an appropriate form) plus the claim that a dependency relation underlies this covariance.[43]

Supervenience alone, and in particular *weak* supervenience, cannot account for the dependence of the mental on the physical that physicalism requires. After all, supervenience without dependence is just co-variation, and co-variation is consistent with any number of mind-body theories. For this reason, Kim writes, "Mind-body supervenience… does not state a solution to the mind-body problem; rather it states the problem itself." (167-168)

Ultimately, Kim draws the following conclusion:

> [T]hese reflections also tell us what needs to be done to upgrade a supervenience claim to the status of a substantive mind-body theory: you must specify the kind of dependence relation that underlies, and accounts for, the mind-body property covariation." (168)

It seems, then, that Blackburn and Kim agree on this much: the (merely) weak supervenience of qualia on physical states is a scenario in need of explanation. The kind of explanation proffered will, of course, have implications for the account as a whole. For this reason, if Kim wishes to affirm the weak supervenience of qualia, then he should give some account whereby the rejection of "mixed worlds" can be rationally justified. (That is, he should *ground* the claim that weak supervenience holds.) How he does so, however, will inevitably have ontological consequences.

---

[43] *Philosophy and Phenomenological Research* 45 (1984): 153-176. Reprinted with permission in Kim, Jaegwon *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993): p. 169 (Page numbers here and elsewhere refer to Kim, 1993.)

To see why, consider the work of David Chalmers. In *The Conscious Mind*, Chalmers endorses a position quite similar to the one put forth by Kim in *Physicalism, or Something Near Enough*. He, too, affirms functional reduction for intentional mental properties and denies the reducibility of qualitative mental states. In light of the failure of strong supervenience, Chalmers maintains that phenomenal properties "naturally supervene" on physical ones; that is, they supervene in the actual world as a consequence of the laws of nature. This may or may not be a understood as a variety of weak supervenience—Chalmers does not tell us whether or not there are any possible worlds in which the laws of nature result in a "mixed word"—but it is clearly *not* a variety of strong supervenience. Kim could embrace a similar picture in defense of the merely weak supervenience of phenomenal properties.

However, it is worth noting that, as a result of this position, Chalmers affirms the following two claims: First, phenomenal properties (qualia) are fundamental features of the world; indeed, they are *just as fundamental* as the most basic of physical properties. Chalmers is unabashedly dualistic, and it's difficult to see how something like the natural supervenience defense could be maintained in the absence of such a strongly dualistic position. Second, among the most basic laws of nature are psychophysical laws, linking phenomenal properties to their physical bases. This is, of course, necessary if the laws of nature are to ensure the supervenience of phenomenal properties on physical ones. It is also a radically dualistic claim, and one that Kim is not likely to find appealing.

We can see, then, that at least one available response to Blackburn's challenge is simply unavailable to an account that professes to be a (mostly) physicalist one. It seems likely that any defense of the merely weak supervenience of qualia on the physical world

will come with some ontological baggage. Until we know which method of response Kim advocates, we cannot really know the full ontological implications of his account.

If, on the other hand, Kim chooses to deny even the weak supervenience of qualia on physical properties, then he will be faced with a different task. Namely, he should provide some explanation as to (a) why his position should be categorized as a variety of physicalism and (b) why qualitative experiences seem to correlate so regularly with physical states. (After all, a good hard kick to the shin *always* results in pain; itch and tickle just *never* seem to follow!) As it stands, it seems to me that the question of whether or not qualia weakly supervene on Kim's account is a genuinely open one.

In conclusion, a physicalist who wishes to adopt Kim's method of reduction through functional definitions as a means of responding to the mind-body problem must be prepared to respond to the following questions: (i) How is it that we can be aware of, and perhaps have knowledge of, epiphenomenal qualia? (ii) Can differences or similarities among qualia be functionalized? (iii) If so, will this result in full causal efficaciousness for qualia differences and similarities? And, finally, (iv) Can we be justified in believing that qualia weakly supervene on the physical world, and—if so—how? Until these questions have been adequately answered, the mind-body problem remains problematic for physicalism.

§6. The Causal Exclusion Problem for Functionally Reduced Mental Properties

Many of the questions just raised are, so to speak, a matter of "hammering out the details."[44] Suppose, then, we bracket these questions and simply consider the merits and dangers of Kim's account as it has already been presented. What follows for the mind-body problem for physicalism? In what follows, I wish to evaluate Kim's functional reductionism, as currently stated, in light of both the problem of mental causation and the problem of consciousness. I will begin with the former.

At first glance, it is easy to see how functional reductions might aid the physicalist in formulating an answer to the problem of mental causation. As Kim has shown us, functionalized mental properties can have causally efficacious realizers; it is natural to suppose that the mental property itself somehow shares in the causal powers of its realizers. Indeed, in *Mind in the Physical World*, Kim formulates the following "causal inheritance principle:"

> If a second-order property *F* is realized on a given occasion by a first-order property *H*, (that is, if *F* is instantiated on a given occasion in virtue of the fact that one of its realizers, *H*, is instantiated on that occasion), then the causal powers of this particular instance of *F* are identical with (or are a subset of) the causal powers of *H* (or of this instance of *H*).[45]

While Kim no longer affirms the first-order/second-order distinction central to the above principle, he surely holds something like this to be the case for functionalized mental properties and their realizers. Whatever causal powers are had by the physical realizers of

[44] This is not to say that they are unimportant; I take them to be extremely interesting and important questions. Still, because Kim offers a response to the mind-body problem, it is surely worth evaluating that response as it has been presented.

[45] Kim, Jaegwon *Mind in the Physical World.* (Cambridge: MIT, 2000) p54-55

a functionally reduced mental property must somehow be inherited, or shared, by the mental state itself.

The question, of course, is *how*. I do not mean to propose that we seek some mechanism whereby the causal powers are passed from one property to another; rather, what we must seek is some account whereby we are justified in affirming the causal efficacy of functionally defined properties solely on the basis of the causal powers of token, non-functional physical properties. Until we have justified this claim, it is difficult to see how functional reduction could "save" mental causation.

Of course, in order to know whether or not mental causation occurs, we need first to know what mental causation requires. What must be the case for mental causation to count as a feature of our world? How we answer this question will bear significantly on whether or not we deem functional reductions successful with respect to mental causation. Kim, like most contemporary metaphysicians, takes causation to be a relation that obtains between events. Furthermore, unless his position has recently changed, Kim understands events to be the exemplification of a property (P) by an object (O) at a time (*t*), such that O's having P at *t* constitutes event E.[46] One possibility, then, is rather straightforward: mental causation must involve an instance of causation where the cause is an event featuring a *mental* constitutive property. [47]

There may be other ways of capturing the truth of mental causation. Still, this is surely a *sufficient* condition of the truth mental causation. If there are instances of the sort

[46] See, for example, Jaegwon Kim "Events as Property Exemplifications," in M. Brand and D. Walton (eds.), *Action Theory*. (Dordrecht: Reidel, 1976) pp. 159-77

[47] This conception of mental causation is not unlike the one proffered by Kim on behalf of the nonreductive physicalist. There, he defines the causal efficacy of the mental as follows: "Mental properties have causal efficacy—that is, their instantiations can, and do, cause other properties, both mental and physical, to be instantiated." (Kim 2005, 35)

described, then surely there is mental causation. However, on Kim's functionally reductive account, it is not at all clear that instances of this sort are possible.[48] For one thing, it is not entirely clear that reduced mental properties retain their ontological status *as* properties given Kim's ontology. In the past, Kim has favored a sparse account of properties, one that does not allow for genuine second-order or functional properties.[49] If this is still Kim's position, then while there are mental terms and concepts, and there are physical properties that satisfy the functionalized definitions of these terms, there are not, strictly speaking, mental *properties* in Kim's ontology. Even if his position has changed, however, it does not follow that these functional mental properties have the requisite degree of causal relevance necessary for mental causation. After all, not every property exemplified by an event-cause qualifies as the property *in virtue of which* the relevant causal instance obtains.[50]

To see the significance of this distinction, note the following example raised by Alvin Plantinga in "Evolution, Epiphenomenalism, Reductionism:"

> When the soprano hits high C and shatters the champagne glass, it is not by virtue of the content or meaning of the line she sings that the glass is shattered; it is just by virtue of the physical properties of the event in question. Even if the words she sings had a wholly different, and indeed contrary content…the glass would have shattered in the same way.[51]

---

[48] This is, of course, *not* to say that mental causation is impossible on Kim's account. I will discuss this further in what follows.

[49] In *Mind in the Physical World*, he explicitly rejects the existence of second-order properties, and indicates as well that functional properties are not, strictly speaking, properties at all. Instead, he suggests that we understand functional properties as "second-order designators," or "second-order concepts." In *Physicalism, or Something Near Enough*, however, he does continue to speak of a "functional property." See, for example, his discussion of the gene on p.163.

[50] Alvin Plantinga helped me to clarify my thoughts in this area, in particular with respect to the "in virtue of" requirement of causal relevance.

[51] In *Philosophy and Phenomenological Research*, 68 (2004): 602-619 (p.604) For the original example, see Fred Dretske, "Reasons and Causes," *Philosophical Perspectives* 3 : 1-15. 1989 (pp.1)

In the above example, Plantinga is referring to the difficulty that a physicalist will have in explaining the causal relevance of the *semantic content* of a mental state. If we prefer, we can restate this challenge in terms of properties.[52] Either way, the point remains: not every feature exemplified by a causally efficacious event is itself causally efficacious.

If a mental property is to be deemed causally efficacious, then there must be at least one instance in which an effect is brought about *in virtue of* the fact that the mental property is present; it is not enough that the property be exemplified by the event that causes the effect. If functional reduction is to save mental causation, then functional mental properties must be able to serve as the constitutive property of a causally efficacious event.

Perhaps this can be accommodated on Kim's account, but if so it is not clear how. After all, if Kim's near-physicalism is correct, then every instance of causation is an instance of causation by a token, purely physical event—that is, an event featuring a particular physical constitutive property. Furthermore, given that the realization relation is not an identity relation, a token physical property will not, as a rule, be identical with any functionalized mental property.[53] In light of this, it seems we can formulate a causal exclusion argument against the causal efficacy of functionalized mental properties—an argument that I have modeled upon Kim's own "Supervenience Argument" against nonreductive physicalism.[54]

---

[52] Suppose the soprano's solo was an instantiation of the property "being a love song." From the fact that the glass was broken by a love song, we certainly may not infer that it was because the song was a love song that the glass was broken. Had the words been of a more hate-laden variety, the results would not have differed.

[53] It may be possible for a physical property to be identical with a functionalized one—if, for example, (a) the functional property had only one realizer and (b) we affirm coextension as the criterion of identity for properties. Still, even if possible, this would surely be the exception.

[54] Kim offers a series of formulations of this argument in (Kim 2005, pp.32-45.)

In formulating his "Supervenience Argument," Kim endorses the following two principles:

> *Exclusion*. No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination. (42)
>
> *Closure*. If a physical event has a cause that occurs at *t*, it has a physical cause that occurs at *t*. (43)

In light of these principles, assume, for *reductio*, that a functionalized mental property M is the cause of some physical event E. (That is, assume that M is the constitutive property of the event that causes E.) According to the functionally reductive picture, M is a cause of E because one of its physical realizers, R, is a cause of E.[55] But then there is a physical event of which R, and not M, is the constitutive property, and that event is *also* the cause of E. Yet R and M are not identical; R *realizes* but *is not* M. Thus, the events featuring R and M are not identical; they have different constitutive properties and, as such, are distinct events. It follows, then, that either E is genuinely overdetermined, or one of the purported causes of E is not the cause of E after all.

Suppose that the physical event featuring R is not actually a cause of E. By closure, we know that there is *some* physical cause of E. By Kim's functionally reductive picture, we know that this cause is not identical to the mental one—its constitutive property is, at most, a *realizer* of M. We are, therefore, back where we started; the exclusion problem cannot be avoided by rejecting the causal relevance of the physical realizer in question. Instead, either the purported mental cause must be rejected, or we must conclude that this is a case of genuine overdetermination. However, because we

[55] Just as Kim's "Supervenience Argument" has multiple possible formulations, we can formulate this argument without this premise. In its place, we need only invoke closure to tell us that there is *some* purely physical cause of E. Because the functionalized mental property is not identical with *any* particular physical property, the argument will go through just the same.

began with the mere assumption that *some* functionalized mental property was causally efficacious, it follows that *no* such property can be uniquely causally efficacious. Either functionalized mental properties are systematically overdetermined, or they are epiphenomenal. In either case, it appears that mental causation—insofar as it requires causation by a mental property—cannot be affirmed.[56][57]

It seems, then, that the causal exclusion problem poses a real challenge for Kim's functional mental properties. Upon further reflection, this should not be surprising. After all, for most (if not all) multiply realized functional properties, their physical realizers will differ with respect to their causal powers. Let F be a functional property and R1 and R2 its (only) two realizers. While R1 and R2 must play the same causal role (C), they need not do so in precisely the same manner. Suppose that they do not. Instead, R1 plays C by causing physical event E1, and R2 does the same by causing E2. Now suppose that

[56] Here I assume, with Kim, that systematic overdetermination is not a "live option" for genuine causation. (Kim 2005, pp.46-52)

[57] There is, of course, one obvious rejoinder here: namely, that this objection fails to take seriously the realization relation. After all, when we claim that the physical property P1 realizes the functionalized mental property M1, we are claiming that every instance of P1's being instantiated is, in some sense, an instance of M1's being instantiated. But this seems to be just another way of saying that every event featuring P1 as a constitutive property is, in some sense, an event featuring M1 as a constitutive property. If the events are identical, then the above argument does not go through. The two purported causes are not rivals, but instead are one and the same. In other words, while P1 might not be type identical with M1, every instance of P1 will be identical at the token level with an instance of M1. If this is enough to ensure identity of events, then perhaps the causal exclusion argument against Kim's functional reduction can be avoided. That said, this seems an unlikely move for Kim to make for at least three reasons. First, multiple realization prevents us from identifying even the instances of M1 and P1. While every instance of P1 will be an instance of M1, the reverse does not follow; M1 might have any number of physical realizers. Additionally, even if we were to bracket concerns about multiple realization, Kim has famously argued against identifying events that differ with respect to their constitutive properties—even in cases where the same physical state exemplifies both properties. He rejects the claim, for example, that the killing of Caesar by Brutus and the stabbing of Caesar by Brutus were the same event, despite the fact that—in this instance, the stabbing was a killing. Finally, this is precisely the move made by Donald Davidson in his attempt to explain mental causation given anomalous monism. In response, Kim argued that, given anomalous monism, no mental event qua mental event could ever be a cause. As such, token identities, even if available, could not grant causal efficacy to mental events. If Kim was correct in claiming that Davidson's anomalous monism ultimately collapses into epiphenomenalism for mental properties, then Kim's own account—should he adopt the strategy just suggested—must do the same.

F is realized at time *t* by R1 and E1 results. For F to be *the cause* of E1, it must be the case that R1 caused E1 *because* R1 was an instance of F.

But this seems patently not to be the case. After all, if F had occurred in R1's absence, then E1 would not have occurred. Despite being an instance of F, R2 would *not* have caused E1, but would instead have caused E2. More simply stated, given a functional account of mental properties—coupled with the claim that mental properties are multiply realized—it looks as if all the causal *work* is done by the physical realizers. There is simply no room for functional mental properties themselves to be causally efficacious.

I do not purport to have shown that mental causation is impossible given functionally reductive physicalism. After all, I simply offered one construal of what it would take for mental causation to be a real feature of this world and shown why, as far as I can see, Kim's account cannot accommodate such a scenario. There may be better ways of understanding the truth of mental causation, and better ways of understanding the causal role of functionally reduced properties. That being said, it is difficult to see how a functionally reductive account of the mental will be able account for a mental property *qua mental* acting as a cause in the physical world. If it cannot, then perhaps we ought to revisit the claim that functional reduction can save mental causation.

§7. The Problem of (Irreducible) Consciousness

Does functional reduction fare any better with respect to the problem of consciousness? It certainly seems to, but only where functional reductions are available.

We have seen how the presence of a realizer can guarantee the presence of a functionalized mental property.[58] If "pain" just *means* "being in some state that is apt to be caused by tissue damage and apt for causing winces and groans," then any creature in such a state will be in pain. With this in hand, we can both (a) explain why Jones was in pain this morning and (b) accurately note that a creature, the phenomenology of which we have no first hand knowledge, will be, or currently is, in pain. Both the explanatory and the predictive gap can be closed by reduction via functionally defined mental properties.

Of course, as Kim himself notes, despite the handiness of the example this is decidedly *not* what it is for a creature to be in pain. "Pain may be associated with certain causal tasks, but these tasks do not define or constitute pain. Pain as a sensory quale is not a functional property." (169) Where qualitative mental states are concerned, functional definitions—and with them, functional reductions—are simply unavailable. As such, they can be of no help to the physicalist who wishes to close the explanatory or predictive gap with respect to *these* properties. On the contrary, reduction through functional definitions offers little hope for the physicalist that this aspect of the problem of consciousness will ever be resolved; in emphasizing the importance of functionalization for reductive explanation, this account closes the door to the possibility of reductively explaining qualitative mental properties.

In other words, functional-reductive definitions cannot assist the physicalist in responding to the *hard* problem of consciousness—other than by giving the physicalist good reason to believe that the problem is insoluble. In *The Conscious Mind*, David Chalmers warns against the following philosophical maneuver:

[58] Here I assume that there *are* functionalized mental properties. The account would have to be changed a bit if these properties were eliminated.

> Frequently, someone putting forward an explanation of consciousness will start by investing the problem with all the gravity of the problem of phenomenal consciousness, but will end by giving an explanation of some aspect of psychological [or functionalizable] consciousness, such as the ability to introspect. This explanation might be worthwhile in its own right, but one is left with the sense that more has been promised than has been delivered.[59]

In *Physicalism, or Something Near Enough*, Kim is careful not to make this mistake. He explicitly states the failure of functional reductions to account for qualitative mental states, and concedes at the outset that his account must inevitably leave certain questions unanswered.[60] Still, while he is more than forthright about this failure, he ultimately concludes that a functionally reductive account is the best that we can do. He writes, "I feel that the position I have been describing here is a plausible terminus for the mind-body debate." (173)[61] This failure to address the *hard* problem of consciousness is not, according to Kim, a defeater for a functionally reductive physicalist account.

The question, then, is whether or not we ought to find this conclusion satisfactory. After all, apart from the hard problem, it's not clear that there *is* a problem of consciousness; Kim himself identifies the problem of the explanatory gap with the hard problem of consciousness, and his account of the predictive gap is stated entirely in terms of phenomenal states. (93-94) If there is no problem of consciousness beyond the hard problem, then Kim's functional reductive physicalism cannot offer a response to the problem of consciousness.

---

[59] (Chalmers 1996, p.26)

[60] See, for example, p29 and p173.

[61] This is not to say that he considers his formulation of this account to be a plausible terminus for the debate. He notes that there are a variety of questions demanding further treatment, and areas of the account that need to be fleshed-out. In the end, however, he takes a position *of this variety* to be a satisfying end to the mind-body discussion.

Consider the following passage, where Kim presents the problem of the explanatory gap:

> Why does pain arise when the C-fibers are activated..., and not under another neural condition? Why doesn't the sensation of itch or tickle arise from C-fiber activation? Why should any conscious experience arise when C-fibers fire? Why should there be something like consciousness in a world that is ultimately nothing but bits of matter scattered over spacetime regions? (13)[62]

At the end of the day, it's not clear that *any* of these questions can be answered on a functionally reductive account.[63] Furthermore, the physicalist who endorses this account will be wholly incapable of addressing the predictive gap as well. Given complete knowledge of the physiology of a creature, such a physicalist would still "have no idea of the qualitative character of its inner experience." (94) It is true, as Kim notes, that functional reductions could close the gap were they available—and could do in a way that surpasses the resources of rival physicalist accounts. Yet if these reductions are unavailable for phenomenal states, then the point is a moot one. In the face of the problem of consciousness, a functionally reductive account can provide no answers.

If qualitative mental states were few and far between, then perhaps we could simply dismiss this problem as an odd sort of mystery and move one. Yet this is not the case. Instead, qualitative experiences are pervasive; they seem to fill every waking moment of our lives.[64] In light of the vast array of functionally irreducible qualitative mental states, it seems that a person who is interested in solving the problem of

---

[62] See also Kim's treatment of the explanatory gap in Ch.4, on p.93.

[63] Kim does offer a potential answer to the final question, that of the very existence of conscious experiences. However, it relies upon the functionalizability of qualia differences and similarities; Kim himself notes that the possibility of such reduction has not yet been shown. (173) I will discuss the difficulties which arise when attempting to reduce qualia differences and similarities in §6.

[64] As Chalmers notes, many of our mental states seem to have both a functionalizable and a phenomenal aspect to them. (Chalmers 1996, 17)

consciousness—the *hard* problem—should not be content with a functionally reductive physicalist account. Of course, it might turn out that *no* account can make sense of qualia; in that case, functionally reductive physicalism could hardly be to blame for failing in the same way. Until that has been shown, however, this failure does seem to be a significant indictment against this account.

Conclusion

In *Physicalism, or Something Near Enough*, Jaegwon Kim presents a complex, persuasively argued and rigorously defended account of the mind. He clearly demonstrates the usefulness of functional reductions in closing the both the explanatory and the predictive gap, and offers a method of reduction whereby mental causation might be preserved. Unfortunately, functional reductions are simply unavailable for the sorts of mental states with which the explanatory and predictive gap are concerned. Because they are unavailable, they are ultimately of no use when it comes to the problem of consciousness. A functional reductionist of Kim's variety will, unlike the identity theorist, be able to formulate the problem of consciousness; she will not, however, be able to address it. Furthermore, as is inevitably be the case in a project of such magnitude, unanswered questions remain. This is not itself problematic, but it remains true that how Kim answers those questions will be crucial to the overall tone and tenability of this philosophical position.

Kim advocates a functionally reductive, mostly physicalist account of the mind as being the best available response to the mind-body problem for contemporary

physicalism. (Furthermore, because he rejects substance dualism outright, he clearly takes this response to be the best available response to the mind-body problem *simpliciter*.) I have argued that this account faces significant difficulties both with respect to the problem of mental causation and, more evidently, with respect to the problem of consciousness. In light of these difficulties, and in light of the questions which remain unanswered on this account, I now wish to turn my attention to alternative accounts of the mind. I suggest that the mind-body problem is not best addressed from the standpoint of reductive physicalism, and that the arguments raised by Kim against nonreductive accounts should be reconsidered. It is to this task that I will turn my attention in subsequent chapters of this dissertation.

CHAPTER TWO:

SUBSTANCE DUALISM AND THE PAIRING PROBLEM

We saw in chapter one that there are phenomena for which Jaegwon Kim's functionally-reductive ontology cannot account. Kim himself claims that these phenomena are nonphysical, and hence that ontological physicalism is false. In this chapter, I would like to revisit a traditional, though in recent years less popular, view of the mind that emphatically affirms the falsity of ontological physicalism. According to substance dualism, there are two radically different types of substances in the world. On the one hand there are material bodies, governed by and wholly explicable in terms of the laws of physics. On the other hand there are minds, or souls, which are *not* material bodies, and which are generally understood to be self-determining. Most, if not all, substance dualists are also *interactionist* dualists; they believe that the immaterial mind is capable of two-way causal interaction with a physical body.[65]

[65] Examples of present-day interactionist substance dualists include Richard Swinburne, John Foster and Al Plantinga. Dean Zimmerman has claimed that, "most days of the week," he can admit to being a substance dualist as well. (Dean Zimmerman, "Material People" in Michael J. Loux, Dean W. Zimmerman, eds. *The Oxford Handbook of Metaphysics* (Oxford: Oxford University Press, 2003) p.492

The primary objection to interactionist substance dualism stems from the purported difficulty that dualists face with respect to mental causation. In "The Rejection of Immaterial Minds"[66], Kim turns his attention to providing a careful articulation of this problem. As Kim rightly states, objections to the possibility of causation between an immaterial substance and a material one are not often articulated in anything like the careful manner we expect to find in philosophical arguments. Instead, the objector typically references Princess Elisabeth of Bohemia and goes on to note how *odd* it is to suppose that such radically diverse substances might causally interact. One example of this can be found in Anthony Kenny's criticism of Cartesian interactionism. (Here, and throughout this paper, I will follow Kim in using "Cartesian" dualism to refer to the interactionist, substance dualist position that also affirms the essential *nonspatiality* of immaterial minds.) In *Descartes*, Kenny writes:

> On Descartes' principles it is difficult to see how an unextended thinking substance can cause motion in an extended unthinking substance and how the extended unthinking substance can cause sensations in the unextended thinking substance. The properties of the two kinds of substance seem to place them in such diverse categories that it is impossible for them to interact. [67]

According to Kenny, the fact that Cartesian minds and physical objects are of such radically "diverse categories" threatens to undermine the possibility of causal interaction between the two.

Though Kim is of course sympathetic both to the sentiment expressed by Kenny and to the conclusion to which his reasoning leads, he nevertheless notes that "As it stands, it is not much of an argument; rather, it only expresses a vague, inchoate dissatisfaction of the sort that ought to prompt us to look for a real argument." (74) It is to

---

[66] Kim 2005 Ch.2

[67] Anthony Kenny, *Descartes* (New York: Random House, 1968) pp.222-223 Cited in Kim 2005 p.74

this task—that of formulating a "real argument" against Cartesian dualism—that Kim turns in formulating the pairing problem.

In what follows, I will attempt to carefully reconstruct and evaluate Kim's pairing problem for Cartesian dualism. I will begin in §1 with an argument that Kim raises against the Cartesian dualist who would affirm Humean causation in order to account for mental causation. §2 consists of a restatement of the primary pairing problem, which Kim takes to apply to all Cartesian dualists regardless of the account of causation they endorse. In §3, I offer three responses on behalf of the Cartesian dualist. The final and most promising response requires the dualist to abandon her belief in the nonspatiality of the mind, embracing instead a form of interactionist substance dualism which affirms the spatial location of the mind. I suggest that this move is unproblematic and can be shown to be more than adequately motivated. Finally, in §4, I consider the implications of the pairing problem for Kim's own functionally-reductive account. I suggest that there is a "qualia pairing problem," analogous to the problem pertaining to immaterial minds, that calls into question the possibility of the epiphenomenal qualia postulated on Kim's account. In order to address this problem, I maintain that Kim should adopt a methodology similar to one that he rejects on behalf of the dualist. I conclude by evaluating potential objections to the apparent parity between substance dualism and Kim's own account with respect to the pairing problem.

§1. Causal Pairing and Humean Causation

Before constructing the primary formulation of the pairing problem, Kim considers a response to Kenny's more general worry about diverse substances which, if successful, might render even a more careful argument against Cartesian dualism unproblematic. According to Louis Loeb, Descartes did not have a robust, metaphysical account of causation. Instead, he viewed causation as mere constant conjunction.[68] As such, Kim writes "On Loeb's view…the fact that soul and body are of such diverse natures was not, for Descartes, even a presumptive barrier to their entering into the most intimate of causal relations." (75) If events of type F are consistently followed by events of type G, and if causation just *is* constant conjunction, then F-events cause G-events; the type of substance involved is irrelevant.

Might a contemporary Cartesian dualist appeal to this Humean conception of causation in order to explain mental causation? Kim thinks not. The problem, according to Kim, is that a Cartesian mind is not eligible for *any* kind of causal relation. Rejecting a metaphysically robust account of causation in favor of a Humean one can do nothing to address the real problem, which stems from the fact that Cartesian minds are essentially nonspatial.

To demonstrate the problem, Kim formulates a preliminary version of the causal pairing problem.[69]

---

[68] The extent to which this reading of Descartes is historically accurate is an interesting question, but not one that is relevant to the discussion on hand. Even if Loeb is wrong about what Descartes actually thought, he might be right about what a contemporary Cartesian dualist *ought* to think.

[69] This is not what Kim calls *the* pairing problem for Cartesian dualism, but it is sufficiently similar to count as *a* pairing problem. The difficulty in both formulations lies in pairing a purported cause with an effect.

> Suppose that two persons, Smith and Jones, are "psychophysically synchronized," as it were, in such a way that each time Smith's mind wills to raise his hand, Jones's mind also wills to raise his (Jones's) hand, and every time they will to raise their hands, their hands rise…Why is it not the case that Smith's volition causes Jones's hand to go up, and that Jones's volition causes Smith's hand to go up? (76)

If causation just is constant conjunction, then Smith's volition does indeed cause Smith's hand to go up. However, argues Kim, by that same reasoning, Smith's volition causes *Jones's* hand to go up—and this will not do. We need some way of correctly pairing minds and bodies such that Smith causes his *and only his* hand to go up, and Jones alone is the cause of his own hand's rising. Furthermore, we need some way of doing this that does not presuppose a union between a given mind and body. When we speak of a mind being "united to" a body, according to Kim, we typically mean that "this body is the only material thing that this mind can *directly* affect." (77) An "ownership" or "union" relation between a mind and a body thus presupposes the possibility of direct causal interaction between the two; it cannot be used to explain that very possibility without thereby begging the question.

In order to correctly pair causes with their effects, we need some sort of pairing relation that holds between the cause and the effect and *fails* to hold between the cause and a rival effect for which that cause is not responsible. If minds are to be causes of material bodies, then they must be able to be joined to those bodies uniquely; absent this union, pairing problems like the one just formulated arise. Appealing to constant conjunction in these cases will not do for, as we saw, constant conjunction cannot suffice for distinguishing the actual cause from the psychophysically synchronized purported cause.

This version of the pairing problem is aimed uniquely at proponents of Humean causation and, as I have said, is not Kim's ultimate formulation. Still, before considering the pairing problem against Cartesian dualism in general, I would like to offer a brief defense of the Humean Cartesian dualist from this line of reasoning. It seems to me that there is a simple and straightforward response to this Humean pairing problem and that is to reject the possibility of the scenario. I do not mean to say that the Humean should reject the possibility of there being two psychophysically synchronized minds acting simultaneously upon bodies. (This would be an unusual scenario, but it's hard to see on what grounds it should be dismissed as *impossible*.) Instead, what the Humean ought to reject is the additional claim that, in cases like the one above, the causal instances can be parsed in the way Kim says they must be.

If we take the Humean conception of causation seriously, then we must say that Smith's volition causes both his hand *and* Jones's hand to rise. Smith's volition is constantly conjoined with Jones's hand rising; as such, Smith's volition *causes* Jones's hand to rise. To say otherwise is simply to insist upon an account of causation that involves something over and above constant conjunction. For this reason, it will not do to insist that, by stipulation, Jones is *not* the cause of Smith's hand rising, though his volition is perfectly synchronized with a distinct volition that *is* the cause. Kim could of course stipulate the possibility of some such scenario, but then it's hard to see why any Humean would accept it. If the Humean grants that, in the example above, Smith might cause his hand to rise and fail to cause Jones's hand to rise, then she has granted that constant conjunction is not always sufficient for causation. If, instead, she denies that this is a coherent possibility, then there is no pairing problem for the Humean Cartesian dualist.

It is worth noting that, in responding to the Humean pairing problem, no mention of the immateriality of the mind was needed. Indeed, we can formulate a pairing problem along precisely the same lines against Humean physicalists as well—provided, that is, that constant conjunction is understood *temporally*, without an additional requirement of spatial contiguity.[70] Suppose light switch A and light switch B are physically synchronized, such that whenever A is turned on, B is as well, and every time they are turned on their respective light bulbs light up. What makes it the case that switch A is the cause of light A's turning on, and not light B's? If causation just is constant conjunction, then *nothing* makes that the case; switch A is just as much a cause of light B's going on as it is of light A. Of course, one could argue that there is a further explanation available here—one that is not available in the case of the Cartesian minds. There is, after all, presumably a chord running from switch A to light bulb A, and not from switch A to light bulb B. Can't this physical connection appropriately pair the switches with their respective light bulbs, and solve the pairing problem?

It can, but only by altering the conception of causation with which we were originally working. As I said, if causation is constant conjunction *combined with* spatial contiguity, then there is a disanalogy here. However, it seems unlikely that a Cartesian dualist would require conjunction *and* contiguity for causation, given that Cartesian minds are said to be outside of space entirely. To adopt such a conception of causation would be tantamount to simply precluding minds from causal relations. Surely this is not what Loeb had in mind when he suggested that, for Descartes, causation was nothing more than "brute regularity" or constant conjunction. (Kim 75) If, instead, causation is

understood to be constant (temporal) conjunction, then the pairing problem is not a problem for Humean Cartesian dualists.[71]

If a contemporary Cartesian dualist embraces constant conjunction as a full account of causation, can she avoid Kim's pairing problem? I think she can. Still, assuming the scenario described in the Humean pairing problem, she may run into some difficulties when trying to make sense of the claim that Smith's mind and Smith's body are united in a way that Smith's mind and Jones's body are not.[72] Is this a real difficulty for Cartesian dualism? After all, if minds and bodies cannot be united, then Cartesian dualism has little to offer as an ontology of mind. That said, it is important to note that the difficulty of pairing minds and bodies really only arises in situations where two minds are said to be psychophysically synchronized. In all other cases, we can simply pair minds and bodies in accordance with observed regularities.

I'm not sure what the dualist should say about a case involving psychophysically synchronized minds and bodies, but I am inclined to think that this is not much of a problem. It seems a perfectly appropriate response for the dualist to simply plead ignorance in such cases, or even to claim that perhaps these minds are somehow united with *two* bodies. In any case, absent any reason to believe that there are such cases, it does not seem to me to be particularly problematic. In general, a Cartesian dualist who

[71] It might be a problem for Humean causation in general, which would in turn be problematic for dualists wishing to affirm Humean causation. That said, this would be the case only if one could give theory-neutral grounds for the claim that cases like the one specified are possible. If it can be shown that cause/effect pairs can be perfectly synchronized without the cause of the one pair thereby counting as a cause of the synchronized effect as well, then causation cannot be mere constant conjunction. (After all, showing that this is possible would amount to showing that there is more to causation than constant conjunction.) It's hard to see how this could be shown without presupposing a thick conception of causation. In any case, the pairing problem, as it stands, offers no unique challenge to Humean Cartesian dualists wishing to affirm mental causation.

[72] There is, of course, something odd about referring to one body as "Smith's body" while disputing the possibility of pairing one and only one body with Smith's mind. I do not mean to presume that one of the bodies really *is* Smith's body; I mean only to designate the two bodies posited in Kim's scenario.

wishes to affirm Humean causation can pair minds and bodies by observing brute regularities between volitions and bodily motions; this will suffice for providing an answer in all cases where it seems reasonable to believe that a clear answer can be had.

Of course, Cartesian dualists need not be Humeans with respect to causation. Kim's principal formulation of the pairing problem is intended to apply to *all* Cartesian dualists, not just those who affirm such a thin conception of causation.[73] In what follows, I will consider the primary argument against mental causation given Cartesian dualism, and offer a defense on behalf of the dualist.

§2. The Pairing Problem Stated

Kim's primary formulation of the pairing problem begins as follows:

> Two guns, A and B, are simultaneously fired, and this results in the simultaneous death of two persons, Adam and Bob. What makes it the case that the firing of A caused Adam's death, and not the other way around? What are the principles that underlie the correct and incorrect *pairing* of cause and effect in a situation like this? (79)

With respect to physical events, such as the one in this example, the solution to the problem is simple. We look for a "pairing relation" that holds between the actual cause and its effect, and that does not hold between the rival cause and the effect.[74] For

[73] More accurately, the primary formulation of the pairing problem is intended to apply to all *non-Humean* Cartesian dualists. The Humean formulation is supposed to show that there is a pairing problem for the Humean, but the primary formulation does not itself pose a challenge to the Humean. By offering both formulations, Kim raises a pairing problem for Cartesian dualists of all causal stripes.

[74] When stating the pairing problem, Kim offers *two* methods of response: one the one hand, we might look for a "continuous causal chain" connecting the actual cause with the effect, and on the other, we might look for a "pairing relation" between the actual cause and the effect. (79) He then notes that the very notion of a "causal chain" presupposes a causal pairing relation between the links of the chain, and so the two options ultimately collapse into the latter of the two. I have simplified things here by mentioning only the second, more successful approach.

example, we might find that A is *pointed at* Adam, and not at Bob, or that A is located at a distance of 100 miles from Bob, but only 100 feet from Adam. Whatever the relation, it will inevitably have to appeal to physical location if it is to solve the pairing problem. This, claims Kim, is why causal pairing is so problematic for immaterial and nonspatial minds.

If Cartesian dualism is correct, then the mind is a wholly non-spatial, immaterial entity capable of interacting with a physical body. Now consider, by way of analogy, a causal pairing scenario involving Cartesian souls:

> There are two souls, A and B, and they perform an identical mental act at time *t*, as a result of which a change occurs in material substance M shortly after *t*. We may suppose that mental actions of the kind involved generally cause physical changes of the sort that happened in M, and, moreover, that in the present case it is soul A's action, not soul B's, that caused the change in M. (80)

In virtue of what could it be true that A's action, and not B's, caused the change in M? Much like we saw in the earlier formulation of the pairing problem, what is needed here is a relation that holds between A and M, but not between B and M, in virtue of which A, and not B, is the cause of the change in M. Furthermore, whatever pairing relation the Cartesian dualist posits must make no appeal to spatial location, and must not implicitly rely upon an established account of causal interaction between the mind and the body; it is the possibility of such interaction that is at issue in the pairing problem, and to assume it would be to beg the question.

What kinds of relationships are available to the Cartesian dualist? Because spatial relations are ruled-out and Cartesian minds are, of course, fundamentally mental, Kim draws the following conclusion:

> Evidently, then, the pairing relation *R* must be some kind of psychological relation. But what could that be? Could *R* be some kind of intentional relation, such as thinking of, picking out, and referring to? (80)

Suppose we assume that an immaterial mind must "pick-out" a material substance prior to causing a change in that substance. Then, the dualist might suggest, the difference between the real cause and the purported cause in the example above is that the former stood in the intentional relation "picking-out" to the effect, and the latter did not. Might this serve as a causal pairing relation?

Kim thinks that it cannot. For we cannot conceive of a thing's "picking out" another substance without first conceiving of its having *perceived* that substance. Furthermore, we cannot understand what it is for an object to perceive a particular substance—as opposed to a distinct yet qualitatively indistinguishable one—without positing some causal relation between the actually observed object and the observer. "Ultimately, these intentional relations must be explained on the basis of causal relations." (81) Intentional relations, then, cannot help the Cartesian dualist; to appeal to an intentional relation in order to solve causal pairing would be to appeal, implicitly, to a *causal* relation in order to solve causal pairing. Because the very possibility of immaterial minds entering into causal relations is what is at issue in the pairing problem, such a response would be obviously question-begging.

Kim further notes that, even if we were able to conceive of intentional relations that do not presuppose causation, this would not suffice to show that intentional relations could serve as pairing relations. In addition, these relations would have to be such that they could individuate intrinsically indiscernible intentional objects. To see why this is the case, it will be helpful to make a slight addition to Kim's formulation of the primary

pairing problem.[75] Suppose that, in addition to the change in M shortly after time *t*, there is a simultaneous change of the very same sort in an intrinsically indiscernible material substance M2. After all, there are *two* minds performing an identical mental act at *t*, and by stipulation acts of that sort typically result in a change in a material body. Kim's formulation leaves open whether one of these mental acts fails to cause such a change or whether, instead, that change occurs in a distinct body not mentioned in the example. Suppose that the latter is true. The result is a situation similar, though not identical, to the one raised against the Humean Cartesian dualist.

Suppose further that the Cartesian dualist wishes to appeal to some intentional relation, R, in order to explain the fact that A's action, and *not* B's action, caused the change in M. If B caused an identical change in M2, then we can assume that R holds between B and M2 as well.[76] The question, then, is this: in virtue of what is it the case that R holds between A and M, but not between A and M2, which is after all intrinsically indiscernible? Can an intentional relation be sufficiently discerning such that it could pick out one material object, but fail to pick-out a distinct yet qualitatively identically one? More importantly, can it do so without implicitly appealing to a causal relation? This, I believe, is where Kim thinks intentional relations are likely to fail. If M and M2 are intrinsically indiscernible, then "picking out" or "thinking about" M will also involve "picking out" or "thinking about" M2.

[75] Kim moves fairly quickly here, and it is not immediately clear why he thinks intentional relations must fail to "suffice for the individuation of intentional objects." (81) I believe that this is a charitable interpretation of what Kim is claiming here, and any misconstrual of his position has been unintentional.

[76] I suppose it's possible that causal pairing might involve a variety of intentional relations, and that some other intentional relation holds between B and M2. Still, there is no harm in assuming that in *this* case, the same relation holds between A and M and B and M2.

Unless intentional relations can be shown to be free of causal presuppositions and capable of such fine-grained distinctions, they will not suffice as a causal pairing relation. What the Cartesian dualist seems to need, Kim concludes, is some kind of non-physical, space-like structure. If there were such a structure, and each immaterial object could be assigned a unique location in this structure, then the dualist could invoke non-physical spatial locations in order to account for causal pairing. That said, we know of no such structure. Kim writes: "I don't think we have any idea what such a framework might look like—what purely psychological relations might generate such a space-like structure. I don't think we have any idea where to begin." (82)

In solving the causal pairing problem for purely physical instances of causation, we appealed to spatial relations. In attempting to solve the pairing problem for instances of causation involving *non-physical* substances, we are unable to do so precisely because no spatio-temporal relations are available to us, and the relations that *are* available do not suffice for the individuation of intentional objects. Kim summarizes the crux of the pairing problem as follows:

> Objects with the same causal powers can differ in the exercise, or manifestation, of their powers, vis-à-vis other objects around them. This calls for a principled way of distinguishing intrinsically identically indiscernible objects in causal situations, and it is plausible that spatial relations provide us with the principal means for doing this. (85)

Without spatial relations, causal pairings cannot be had; causation presupposes location in physical space. This is the moral of the pairing problem.

What follows for Cartesian dualism if Kim is correct about causal pairing? First, and most obviously, Cartesian dualism cannot accommodate mental-to-physical causation. Nonspatial, immaterial minds cannot exert any causal influence upon the material world, and this includes physical human bodies. But this is really only the

beginning of the problem. Cartesian minds would be equally incapable of causally interacting with one another. The pairing problem is not limited to "diverse substances", but instead calls into question the possibility of immaterial minds acting as causes *at all*. The nature of the substance with which a mind is purported to interact is wholly irrelevant; interaction itself is the problem. Finally, a Cartesian mind—on this picture—cannot be the causal byproduct, or epiphenomenal result, of anything physical or nonphysical. If the causal relation is essentially a spatial one—if it presupposes spatial relations—then a mind cannot enter into any causal relations. This holds for *both* relata; a Cartesian mind can neither cause *nor be caused by* anything at all.

§3. Responding to the Pairing Problem

What should the Cartesian dualist say to Kim's causal pairing problem? This argument, if successful, threatens to eviscerate Cartesian dualism; unlike vaguely expressed worries about "diverse substances," this is not the kind of objection that a dualist can simply dismiss. Furthermore, unlike the scenario described in the argument against Humean Cartesian dualism, this scenario really must be possible if Cartesian dualism is true. At the very least, it's difficult to see on what grounds one could reject it as impossible. Surely you and I can successfully will to raise our hands simultaneously *on occasion*; this is really all that the pairing problem requires.

Fortunately, I believe that there are a number of responses available to the interactionist dualist—though not all of them will be available to a *Cartesian* dualist, strictly speaking. There are, to my mind, three distinct lines of response which a

substance dualist might advance. I believe that any of the three could suffice as a response the pairing problem, though the degree to which they can be made plausible varies significantly. Each of the three comes with its own set of difficulties and advantages; I will treat them in turn, beginning with what I take to be the weakest response and ending with the strongest.

Before considering these responses, it is worth noting that a Humean account of causation should be just as effective against the primary formulation of the pairing problem as it was against the version of the argument aimed specifically against Humean Cartesians.[77] (In light of this, I suppose there are really *four* lines of response available to the substance dualist, three of which are available to those wishing to affirm a more metaphysically robust account of causation.) The potential problem for Humean mental causation stemmed from the possibility of a psychophysical synchronization of causes. The primary formulation of the pairing problem makes no reference to such synchronization, and adds no new challenge pertinent to Humean causation. Still, rather than commit all Cartesian dualists to a Humean account of causation, it seems best to consider alternative responses.

§3.1 Finding, or Not finding, a Pairing Relationship

First, a Cartesian dualist might simply deny the claim, implicit in Kim's argument, that causal relations must be transparent. Consider once again the purportedly problematic case of causation by souls A and B:

[77] Thanks to Leopold Stubenberg, who pointed this out in an earlier draft of this paper.

> There are two souls, A and B, and they perform an identical mental act at time *t*, as a result of which a change occurs in material substance M shortly after *t*. We may suppose that mental actions of the kind involved generally cause physical changes of the sort that happened in M, and, moreover, that in the present case it is soul A's action, not soul B's, that caused the change in M. (80)

What the pairing problem demands is some pairing relation in virtue of which A, and not B, can be truly said to be the cause of the change in M. In requiring that we be able to *identify* some such relation, might Kim be slipping from an epistemological worry to a metaphysical conclusion?

There is one clear sense in which he is not doing this. Kim is not claiming that, because we can't see which soul is the cause of the change in M, there must not be any fact of the matter to serve as an answer to that question. Instead, he has carefully articulated certain requirements which a relation must meet if it is to suffice as a causal pairing relation. Then, having considered the relations that he takes to be the best candidates for causal pairing, he concludes that none of the relations available to the dualist can do the work necessary to resolve the problem. Only then does Kim conclude that spatial relations are necessary, and thus that causal relations presuppose spatial location.

Still, there is a move here that the Cartesian dualist can reject. Just because we have considered the relations that we thought were best suited for causal pairing and failed to find any decent candidates, it need not follow that there *are* no candidates. For all we know, there are relations which obtain between immaterial causes and their effects that are simply beyond our epistemic reach. All that is necessary for mental causation given a Cartesian ontology of mind is that causes and effects be related to each other in a sufficiently discerning way. Our ability to articulate *how* they are so related is an additional requirement, and one that need not be deemed necessary by Cartesian dualists.

Furthermore, in light of the fact that it is *mental* causal pairing relations with which we are here concerned, it seems particularly important to consider the difference between the first-person and the third-person perspective.[78] I might be very good at determining which of my mental actions was responsible for causing some physical event, but decidedly less reliable when it comes to ascribing mental causes to the actions of others. If the dualist could show that there are eligible pairing relations that suffice from a first-person perspective, then perhaps she will have done enough to defend Cartesian dualism from the pairing problem. There might remain ambiguous cases in which we are unable to distinguish the true cause from a purported rival cause, but that should not undermine our confidence in *all* instances of mental causation. After all, we need not believe that our methods of ascription are infallible, only that they are in general reliable.[79]

There are, to be sure, some difficulties with this response. Depending upon the degree of inscrutability the dualist ascribes to mental pairing relations, this response threatens to relegate mental causation to the realm of mystery without offering any hope of rescuing it any time soon—or even at all. The Cartesian dualist who offered this line of reasoning could be perfectly justified in claiming that causation by immaterial minds occurs. (That is, if she were independently justified in believing that the mind is immaterial and that mental causation occurs, the pairing problem alone would not

[78] Michael Morris alerted me to the importance of first-person perspective when considering mental relations.

[79] It is worth noting, again, that in those cases in which we could not reliably distinguish a true cause from a rival causal candidate, it would *not* follow that there was no true cause. Our inability to recognize a relation does not guarantee, or even imply, that the relation does not obtain.

undermine that justification.)[80] However, having made the claim that causal paring relations between mental causes and their effects are inscrutable, she would face difficulties when seeking justification for the ascription of any particular effect to a given mental cause. Beyond the mere claim that it occurs, it is difficult to see how such a dualist could have very much to say about mental causation. If, however, the dualist could give an articulation of some pairing relation that suffices for the pairing of causes with their effects *from the first person perspective*, then perhaps the fact that these relations are sometimes inscrutable from a third-person perspective will be less problematic.

§3.2 Causation and Intentionality

The second response available to a Cartesian dualist is similar to the first, though it is perhaps more promising. Rather than appeal to the inscrutability of mental pairing relations, the dualist might instead seek to defend one type of relation considered and rejected by Kim: intentional relations. As we saw, Kim had two reasons for rejecting the possibility of intentional pairing relations. First, he claimed that they must inevitably appeal to perceptual experiences, which in turn rely upon established causal relations. Second, even if they did not, he deemed them unlikely to suffice for the individuation of intrinsically indiscernible intentional objects. I will begin with the latter, which—if true—would render a response to the former unnecessary and unhelpful.

[80] I will say more about the effect of the pairing problem on independent justification for one's ontology of mind in subsequent sections of this chapter.

It seems to me that the best way for a Cartesian dualist to respond to this objection is to take it as a challenge, rather than a defeater. After all, Kim has not shown, nor does he say, that intentional relations are *unable* do the individuation work required of pairing relations, only that it seems *unlikely* that they could do so. (81) The primary reason he gives for this belief stems from the fact that it is difficult to conceive of a fully fleshed-out coordinate system, analogous to space-time, comprised only of intentional or psychological relations. (82) It is not immediately clear to me that such a complex coordinate system would be necessary for causal pairing. If this follows from the pairing problem, I do not yet see how it follows. What does follow, and what is required is that an intentional relation be able to pick-out one particular intentional object, without thereby picking-out distinct yet intrinsically indiscernible ones.

At first glance, this certainly seems possible. The intentional attitudes of which I am aware seem to be capable of making fine-grained distinctions. When I think of my favorite tree by the Setauket Mill Pond, I am quite confident that it is *that* tree—and not some intrinsically indiscernible one—about which I am thinking. Of course, I have had perceptual experiences of the tree by the Setauket Mill Pond. These perceptual experiences play an important role in my ability to pick-out *that* tree, in all its particularity. Indeed, according to Kim, perceptual experience is necessary for the sorts of intentional relations being considered, at least in cases when the purported intentional object is a "concrete thing outside us." (81)[81] This is, of course, Kim's first objection to intentional pairing relations.

[81] It might be possible for the Cartesian dualist to argue that the intentional relations obtaining between a soul and a body are more like those obtaining between a person, however understood, and her own beliefs. To do so, however, she would have to affirm some sort of unity between the soul and the body without thereby presupposing causal interaction between the two. I am inclined to agree with Kim in thinking this to be a difficult task.

Is it possible, then, to address the second objection—the claim that intentional relations are insufficiently discerning—without immediately being confronted with the first? There is one way that this might be done. I might simply stipulate that some non-causal intentional relation had been found, call it R, and then consider whether R could suffice for causal pairing. In doing so, I would be assuming success with respect to the first objection in order to assess the strength of the second objection. There is a sense in which this seems to be a fair response. After all, the second objection claims that *even if* a thoroughly non-causal intentional relation could be found, it would be insufficiently discerning. Why shouldn't the dualist start with the assumption that the first criterion has been met, and attempt to see if the troubling conclusion follows?

This seems to me to be an unpromising approach. It's hard to see how we could know much about R's ability to individuate its relata without knowing more about what exactly R is. It is, of course, equally difficult to see how we could know that R would fail to suffice as a pairing relation without this additional information. Furthermore, Kim gives no distinct argument for the claim that non-causal intentional relations will be insufficiently discerning. It seems, then, that Kim's second objection really has to be considered in conjunction with the first. In fact, I suggest that there are not really two objections here, properly understood, but one. The problem is not that some intentional relations will be precluded because they presuppose causation and others will turn-out to be incapable of the requisite degree of specificity. Rather, *because* they are not causal relations, intentional relations which do not presuppose causation will fail to distinguish between indiscernible entities. The crux of Kim's claim here can, I think, be summarized as follows: If an intentional relation is to be sufficiently discerning, it *must* ultimately appeal to a causal relation.

If we understand Kim's objections in this way, then the task of the dualist seems clear. She must find an intentional relation that succeeds in individuating indiscernible entities without ultimately collapsing into a causal relation. For this to be the case, according to Kim, I must also not presuppose perception, for perception itself depends upon causation. What is needed, then, is an example in which I stand in an intentional relation to one particular object, having had no perceptual experience of that object. Is this a possible scenario?

The answer to that question seems to depend upon how we understand "perceptual experience." Is indirect perceptual experience a problem, or only direct experience? If the latter, then surely this criterion can be met. Every time I reflect upon President Bush I succeed in individuating a single person of whom I have had no direct perceptual experience. Surely, then, indirect experience must also be a problem. But what exactly counts as indirect perceptual experience? Any experience that I have had of Napoleon, for example, is so far removed from Napoleon himself as to be essentially untraceable. From the fact that I can think about Napoleon, do I really have to conclude that I have had some perceptual experience of him? If not, must I instead conclude that, despite what I think, I can't really think of Napoleon himself; that, at best, I can only pick out some small class of short, aggressive, French military leaders?

It is important that we understand this claim—that sufficiently discerning intentional relations must appeal to perceptual ones—in light of the additional claim that all perceptual relations are causal ones. If both claims are true, then it must be that either

Napoleon is somehow *causally* responsible for any thoughts that I have of him, or I cannot really have thoughts of Napoleon himself after all.[82]

On the other hand, if either claim is false, then the dualist could have a response to the pairing problem. If a sufficiently discerning intentional relation could be found that does *not* rely upon any perceptual experience, then this relation could serve as a pairing relation. Alternatively, if perception were understood non-causally, then all that would be needed would be an intentional relation that is not causally laden; perception would no longer be problematic. Of the two options, I find the second to be the more promising one. Even if we can find an intentional relation that makes no appeal to perception, it seems likely that many intentional relations—particularly the ones that are likely to be most relevant to mental causation—will involve direct perception of the intentional object. The dualist who chooses to defend intentional pairing relations, then, ought to appeal to a non-causal account of perception.

Before attempting to undermine the centrality of causation to perception, it will be helpful to see why Kim affirms it. In support of this position, Kim formulates what we might call a *perceptual* pairing problem:

> What is it for me to perceive this tree, not another tree which is hidden behind it and which is qualitatively indistinguishable from it? The only credible answer we have is the familiar causal account, according to which the tree that I perceive is the one that is causing my perceptual experience as of a tree, and I do not see the hidden tree because it bears no causal relation to my perceptual experience. (81)

By now this line of reasoning ought to be familiar. Whatever relation perception is, it must be the kind of relation that is capable of distinguishing qualitatively indiscernible objects. A causal relation will be capable of doing just that, and as far as we know, no

[82] This is, of course, consistent with Kripke's causal account of reference. See Saul Kripke, *Naming and Necessity.* (Cambridge: Harvard University Press, 1980)

other intentional relation is so capable. (Presumably, the causal relation to which Kim here refers must ultimately appeal to spatial location as well.)

Of course, the claim that only causal relations can serve as pairing relations has not yet been established. The pairing problem shows us the causal relation must be able to distinguish intrinsically indiscernible objects. What has not yet been shown is this additional claim, that any relation which is capable of making such distinctions must be a causal one. For this reason, if the dualist can offer an account of perception that does not appeal to causation and that does serve as a response to the perceptual pairing problem, then she can reject the claim that perception presupposes causation and, with it, the claim that intentional pairing relations cannot be had.

Here I think the dualist has options. William Alston, for example, defends a "Theory of Appearing", which makes no reference to any causal relation obtaining between the perceived object and the perceiver. In "Perception and Representation", he writes:

> To have a certain kind of perceptual experience is for an object to appear to the subject as such and such, to look large or moving or droopy or like a trillium.' The appearing object is part of what makes the experience what it is. One could not have just that experience without just that object's appearing to the subject as it does.[83]

If this theory can be defended, then it will serve as a noncausal response to the perceptual pairing problem. On this account, what makes it the case that I am perceiving this tree, and not another, qualitatively indistinguishable one, is the fact that *this* three constitutes part of my perceptual experience, and the other tree does not.

---

[83] Alston, William "Perception and Representation" *Philosophy and Phenomenological Research* Vol. LXX, No. 2, March 2005 p.257 Leopold Stubenberg alerted me to this account, and I owe this reference to his helpful comments.

With the Theory of Appearing in hand, the dualist could thus formulate a response to Kim's pairing problem. First, she could invoke intentional relations—for example, "picking-out"—to explain causal pairing. What makes it the case that one mind, and not another, was the cause of a given change in an object would thus be the fact that the former, and not the latter, stood in the right sort of intentional relation to that object. While it seems likely that intentional relations presuppose perceptual ones, this need not be a problem. If the perceptual relation is a noncausal one, as we saw in Alston's account, then the intentional relation can be shown to be free of causal presuppositions. Furthermore, because of the role played by the perceived object in the Appearing relation, this relation will also suffice to individuate among qualitatively indiscernible objects. The dualist, then, can address the pairing problem without presupposing causation by an immaterial mind.

There is, however, one remaining concern. Alston's account makes no reference to a causal relation obtaining between the perceiver and the perceived object, and the appearing relation seems to be of the sort that is capable of distinguishing indiscernible objects. However, if it turns out to the be the case that this relation itself presupposes spatial location, then the dualist is right back where she started. For even if the appearing relation is not a causal one, if it requires that the relata be spatially located, then a mind without location would be incapable of entering into any such relation.

I believe that the dualist can resist this move, but that doing so invites difficulties. From the fact that material bodies need to be in relatively close proximity with one another in order to (directly) perceive each other, it need not follow that the same holds for immaterial minds. There is nothing incoherent in supposing that minds are able to "turn their attention" to any object at all, regardless of where that object is located.

However, it certainly seems as if *our* minds are restricted by location. As I sit here in the library, I can perceive the book that is currently sitting next to my laptop; try as I might, I cannot currently perceive the books on my bookshelf at home. It seems, then, that spatial location is somehow intimately involved with our perceptual experience; it seems, even, that location is a necessary feature of perception.

That said, it is the location of the *body*, and not of the mind that is so central to perception. The perceiving mind, it seems, must somehow be related to a spatially located body in a way that explains the constraints that the body's location places on the mind's perceptual abilities. The dualist who chooses to affirm a noncausal account of perception coupled with intentional pairing relations must find a way to articulate this relationship, and must do so in a way that does not implicitly appeal to causal interaction between the mind and the body.

§3.3 Abandoning *Cartesian* Dualism

The third and final response available to the substance dualist is quite different from the first two.[84] Rather than deny the necessity of spatial location for causation, the dualist might instead accept this conclusion of the pairing problem and, in response, choose to locate immaterial minds in space. The Cartesian dualist would no longer qualify as "Cartesian" by the standards we have been using, but this hardly seems problematic in and of itself. This seems to me to be the most promising response for the substance

[84] There may, of course, be additional responses as well; I mean only that this is the third response that I know of that the dualist might take.

dualist to take, and the one that comes with the fewest costs. It is also the response to which Kim devotes the most attention, so it will be helpful to treat each of his objections in turn.

First, Kim notes, if minds are to be located, then we need a place to put them. More importantly, we need "a motivated way" of determining where in space these immaterial minds should go. (88) Given that they are in space, *where* in space are they? He writes,

> It would beg the question to locate my soul where my body, or brain, is on the ground that my soul and my body are in direct causal interaction with each other; the reason is that the possibility of such interaction is what is at issue and we are considering the localizability of souls in order to make mind-body causation possible. (89)

Despite this claim, I wish to suggest that would be not at all question begging for the substance dualist to locate the mind in the body—or anywhere at all—on these grounds. After all, the question with which the dualist is confronted is the following: Given the necessity of spatial relations for causation, how is it possible for an immaterial mind and a body to interact? If the dualist responds by saying "It is possible because the mind is located where the body is located", then she has offered an *answer* to the question, she has not begged it. Notice that the dualist has not said "It is possible because the mind and the body are causally connected." The latter would be question begging; the former is not.

In fact, even if the dualist were to go on to say "I know that the mind is located because the mind and the body interact causally," she would *still* not be guilty of begging the question. Indeed, it seems that this is precisely how a substance dualist who is both committed to the immateriality of the mind and convinced by the pairing problem should respond. If I believe that the mind is both immaterial and causally efficacious, and I come to believe that causally efficacious entities must be spatially located, then I ought to

conclude that the mind is spatially located.[85] If it turns out that the location must be *in* the body if it is to be of any help, then I ought to conclude that the mind is located in the body. This is a perfectly respectable philosophical move, and the dualist can employ this response without begging any important questions. Perhaps Kim's real worry here is that the dualist could not be *properly motivated* to locate the mind in the body, rather than to conclude that the mind is material, when faced with the pairing problem. I will return to this question when responding to the next objection. For now, it will suffice to note that the question-begging worry is itself not a problem for the substance dualist.

In his second objection to the location of immaterial minds in space, Kim notes that location alone will not suffice as a response to the pairing problem. In addition, something like the "impenetrability of matter" must hold for minds as well. Kim writes,

> It must be the case that no more than one soul can occupy a single spatial point; for otherwise spatial relations would not suffice to uniquely identify each soul in relation to other souls in space. (89)

In light of this, it seems it is not merely the fact that physical objects can be located in space that allows for them to be causally paired. Instead, their location must be *coupled with* the fact that they are the sole occupants of that location. If immaterial minds are to be eligible for causal pairing, then they too must occupy their locations uniquely. (At least, they must be the only *mind* at that location; it need not follow that they cannot share space with a material object.)

[85] It is important to note that there is nothing internally incoherent about locating an immaterial mind in space. (It would be incoherent to posit the location of a Cartesian mind, as a Cartesian mind is purported to be essentially nonspatial.) Absent some argument demonstrating the materiality of spatially located objects, the dualist remains free to locate the mind in space on the grounds being considered here.

Of course, as Kim concedes, the dualist could simply posit a principle whereby no two souls could ever occupy the same region of space at the same time. In response to this possibility, Kim writes:

> To solve the pairing problem for souls by placing them in space we need such a principle, but that is not a reason for thinking that the principle is true. We cannot wish it into truth—we need independent reasons and evidence. (90)

The sentiments expressed here are, I believe, at the heart of both this objection and the preceding one. It is not question begging for the substance dualist to posit a principle of spatial exclusion in response to the pairing problem, just as, I suggest, it is not question begging for the dualist to locate the mind in space for the same reasons. However, if the dualist is to be rationally justified in making either of these claims, she must have some "independent reasons and evidence" in favor of their truth.

I think that Kim is correct about this. However, I think that most reflective substance dualists could provide such reasons. To see that this is the case, it is important to note that one's evidence in favor of a given ontology of mind need not be limited to that which is directly relevant to the pairing problem. Consider the following criticism by Kim:

> It may be that one's dualist commitments dictate certain answer to these questions. But that would hardly show that they are the 'correct' answers. When we think of the myriad questions and puzzles that arise from locating souls in physical space, it is difficult to escape the conclusion that whatever answers might be offered to these questions would likely look ad hoc and fail to convince. (90)

It is certainly the case that, given only the questions raised by the pairing problem, a physical mind seems less problematic, and more plausible, than an immaterial one. That said, causation is but one feature of the world. Mental causation is an important problem for any account of the mind, but it is far from being the *only* problem.

The substance dualist must be able to show that the immaterial mind is compatible with mental causation if she is to defend substance dualism. She need *not* show that the immateriality of the mind is the most plausible position in light of mental causation. There are a host of other concerns—the inherent subjectivity and unity of consciousness; the rational, rather than mechanistic, nature of reasoning; the importance of content to mental states; the appearance of genuine libertarian freedom and deliberation; questions of personal identity; the possibility of life after death—all of which, though not necessarily relevant to a discussion of causal pairing, are directly relevant to questions of the ontology of the mind.

If, in the face of the pairing problem, the substance dualist is compelled to locate the mind in space, the extent to which this move is motivated must be judged in relation to all of her evidence in favor of substance dualism, not just that which pertains to the pairing problem. Kim is right to claim that "we need independent reasons and evidence" in support of the claims that the mind is spatially located and subject to spatial exclusion. It does not follow that this evidence must be limited to that which is directly relevant to causal pairing. Instead, the dualist should feel free to use the conclusion of the pairing problem, coupled with her previously established reasons in favor of the immateriality and causal efficaciousness of the mind, to conclude that the mind must have a unique spatial location. Such a move does not beg any important questions, and can be adequately motivated and rationally justified.

Kim raises two additional objections to the location of immaterial minds in space, though his treatment of both is quite brief. First, he poses the following question:

> If souls are subject to spatial exclusion, in addition to the fact that the exercise of their causal powers are constrained by spatial relations, why aren't souls just material objects, albeit of a very special, and strange, kind? (90)

I will discuss the question of what it is for something to qualify as a material, or physical, object in a subsequent chapter. For now, I think the following response will suffice. If the physicalist is willing to radically revise her understanding of the physical world in a way that allows for the sorts of things that we have been calling immaterial minds—complete with libertarian freedom, subjective unity and the possibility of surviving bodily death—then I see no reason why the dualist should insist upon a duality of substance. If, instead, Kim is suggesting that the dualist simply concede *physicalism*, as it has been traditionally understood, then I see no reason why the dualist should come to this conclusion in light of the spatial location of minds.[86] After all, it is not as if the notion of a spatially extended immaterial substance has been shown to be *incoherent*. Once again, because the dualist's commitments to the immateriality of mind are rooted in considerations which go beyond those raised in the discussion of the pairing problem, she can be perfectly well motivated to maintain these commitments—even after having accepted spatial location.

Kim's final criticism centers upon the location of the mind at a single geometric point. He writes,

> If a soul, all of it, is at a geometric point, it is puzzling how it could have enough structure to account for all the marvelous causal work it is supposed to perform and how one might explain the differences between souls in regard to their causal powers. (90)

It is important to note that this objection only holds for those dualists who choose to locate the mind at a single point, rather than as a spatially extended substance. I believe

[86] William Hasker responds in this way to a similar criticism when defending his "emergent dualism" account of the mind. To those who would claim that Hasker's emergent mind is physical, he writes "If philosophers are prepared to stretch the meaning of 'physical' to encompass everything that has been said here about the field of consciousness, then so be it. What is *not* acceptable, however, is for someone to take the claim, thus arrived at, that 'the mind is physical' and use it as a premise from which to infer characteristics of the conscious mind that are contrary to the ones postulated in this chapter." William Hasker, *The Emergent Self*. (Ithaca and London: Cornell University Press, 1999), p.201.

that both options remain available to the dualist, but the scope of this discussion does not allow for an adequate treatment of the latter possibility.[87] Suppose that the dualist does locate the mind at a single, extensionless point. Does it follow that the soul must be cripplingly simple, in a way that will not allow for mental causation?

I don't see how it does. From the fact that a soul *has* spatial properties, it does not follow that *all* of the soul's properties are spatial. Kim of course recognizes this, noting the possible appeal to a "mental structure" to account for the requisite degree of complexity. In response, he offers a series of hypothetical questions about this mental structure. This objection does not seem to be central to Kim's arguments against the location of mind, and he devotes only a few lines to treating it. Perhaps there is something more to the objection than is immediately evident, but as I understand it, it does not seem to be a source of real trouble for the dualist. When the dualist claims that the mind is immaterial, she is claiming that it is a substance that can neither be understood nor explained in purely physical terms. It is not surprising, then, to learn that spatial location cannot suffice to explain the mind's complexity. The further claim, that location at a point is *incompatible* with having a complex mental structure, has not been shown.

[87] I will say this much. To my knowledge, the primary reason against positing extended immaterial substances has, historically, been the belief that with spatial extension comes corruptibility. If the soul is about the size of my brain, then I should be able to split it in half, just as I would a material object. I don't see that this follows from spatial extension, though I realize that this is a claim that requires a greater defense than I can here provide. In thinking about these things, however, I believe that it is important that we reflect on the kinds of things that physics now posits. A field, for example, cannot be "broken" simply by running a knife through it, though a field is surely extended in space. From the fact that an entity is extended in space, we should be hesitant to conclude that it thereby has all of the properties of a solid material body. Absent this conclusion, I see no reason to conclude that the mind, if located, must be at a single extensionless point.

§4. The Qualia Pairing Problem

Kim's pairing problem constitutes a direct, carefully articulated challenge to substance dualism. Absent some response to this problem, it is difficult to see how one could affirm the causal relevance of an immaterial mind. Fortunately, as I have argued, the dualist has a number of responses available. She could affirm a Humean account of causation, insist upon the inscrutability of mental pairing relations, attempt to defend intentional pairing relations, or conclude that immaterial minds must be spatially located. Each of these responses comes with its own set of difficulties, though it seems that the final option is likely to be the most promising. The pairing problem does demand a response, but this is a demand that the substance dualist can meet.

In what follows, I would like to consider an interesting, unintended consequence of Kim's pairing problem. As we saw in §3.3, the primary difficulty faced by the dualist wishing to affirm spatial location was one of motivation. I argued that the dualist could be adequately motivated to locate minds in space, provided that we consider the broad basis of her philosophical commitments. I will now argue that Kim must employ a similar line of reasoning in response to what I will call the "qualia pairing problem." Just as the dualist must appeal to her larger theoretical commitments in order to motivate the location of minds in space, so too must Kim appeal to his theoretical commitments in order to motivate the location of qualia in space.

§4.1 The Problem Stated

As we saw in Chapter One, Kim advocates the functional reduction of the mental to the physical in order to account for mental causation. However, despite the success of functional reductions with respect to cognitive and intentional mental properties, Kim concludes that qualia cannot be so reduced. The metaphysical possibility of a qualia inversion demonstrates the fact that qualia cannot be given functional definitions; any definition in terms of a causal role must inevitably fail to capture the intrinsic properties of qualia. Because intrinsic properties are essential to qualia—as we learn from Kripke, a pain must of necessity *feel* like a pain—functional definitions cannot be employed in the case of qualia.[88] In light of this difficulty, Kim concludes that qualia must not be physical. He writes, "Qualia, therefore, are the 'mental residue' that cannot be accommodated within the physical domain. This means that global physicalism is untenable." (170)

For a physicalist, the failure of global physicalism is obviously a difficult and troubling position upon which to arrive. Kim defends this position by noting that qualia, because they are irreducible, must also be epiphenomenal. "The mental residue, insofar as it resists physical reduction, remains epiphenomenal. It has no place in the causal structure of the world and no role in its evolution and development."(171) By insisting upon the epiphenomenal nature of qualia, Kim is able to avoid difficulties with respect to the causal closure of the physical and the threat of systematic overdetermination.[89]

That said, it seems unlikely that Kim really intends to affirm the claim that qualia have "no place in the causal structure of the world." After all, an epiphenomenal result is

[88] (Kripke, 1980) Lecture 3.

[89] Indeed, it is *because* of closure and exclusion that Kim concludes that qualia must be epiphenomenal. I will discuss this "causal exclusion problem" in the remaining two chapters.

one that was *caused*; qualia must, therefore, be able to enter into causal relations. Perhaps they are unable to serve as the cause in any causal instance, but they must be able to serve as an effect. To see why this is the case, consider a world in which qualia could not be caused. In this world, it would not only be the case that the feeling of pain could not cause me to wince—a startling discovery to be sure, but a coherent one given Kim's metaphysics—but it must also be the case that a kick to the shin could not cause me to feel pain. Likewise, my feelings of hunger could not be caused by my lack of eating, and I could not cause the feeling of relief from hunger by eating. If qualia were truly incapable of entering into causal relations at all, the world would be an odd place indeed. If *our* world is a world like that, then there is much to be explained. In the actual world, qualia correspond regularly with physical events; a kick to the shin, if it's hard enough, will always cause the feeling of pain. It's hard to see how this correlation could be explained without allowing qualia to be caused.[90]

Qualia, then, though epiphenomenal, must not be wholly causally irrelevant. The question, in light of the pairing problem, is whether this position is a tenable one. Can wholly irreducible qualia be the epiphenomenal result of causation? If spatial location can be had only by physical entities, then it's not clear that they can. To see why, consider the following scenario:

> There are two bodies, A and B, and they undergo an identical physical change at time *t*, as a result of which a qualitative change Q occurs shortly after *t*. We may suppose that physical changes of the kind involved generally cause qualitative changes of the sort that happened at *t*, and, moreover, that in the present case it is the change in A, not B, that caused Q.

[90] I do not mean to suggest that a causal explanation is the *only* available explanation of this correlation. Still, if qualia cannot be caused, then a vast number of what appear to be causal instances must not be causal instances. Surely if this result can be avoided, it should be avoided.

This is, of course, a minor restatement of the primary formulation of Kim's pairing problem.[91] The question, one with which we are now quite familiar, is how this is to be understood. In virtue of what is it the case that A, and not B, is the cause of Q? What kind of pairing relation might hold between a material body A and the occurrence of a wholly nonphysical quale?

Like the substance dualist, Kim has a variety of responses available here. He might affirm Humean causation, insist upon the inscrutability of qualia causation, defend an intentional relation which could pair physical causes with qualitative effects, or locate qualia in space. Also like the substance dualist, indeed to a far greater extent than the substance dualist, the final option seems the most promising.

There are, however, initial difficulties that Kim will face should he choose to locate qualia in space. The claim that spatial location is a physical property, had by physical objects, is one that Kim makes repeatedly in *Physicalism, or Something Near Enough*.[92] Consider the following passage, taken from the conclusion of this work:

> Causality requires a domain with a space-like structure—that is, a 'space' within which objects and events can be identified by their 'locations'—and, as far as we know, *the domain of physical objects is the only domain with a structure of that kind*. (151, my emphasis)

Furthermore, he repeatedly asserts the radical nonphysicality of qualia. Qualia, he writes, "cannot be accommodated within" and "stay outside of the physical domain." (171,173) If qualia must remain outside of the physical domain, and the physical domain is the only one with a spatial structure, then it seems qualia cannot be spatially located.

[91] I have not given a location of the qualitative change because it is the ability to assign such a location to a quale that is at issue.

[92] See, for example, pp72, 80, and 151.

At the same time, given only the considerations raised by the pairing problem, it is difficult to see *why* Kim would choose to locate nonphysical qualia. It seems instead that he should simply concede the truth of global physicalism and reject nonphysical qualia. If causation requires spatial location, and location—as far as we know—is had by physical objects, then the most plausible conclusion seems to be that the causal relation is available only to physical objects. Any attempt to defend the location of epiphenomenal qualia, then, *from the perspective generated by the pairing problem*, must seem ad hoc and unmotivated.

Despite these initial appearances, however, Kim can be both justified and motivated in choosing to locate qualia in space. Unless the notion of a located quale can be shown to be incoherent, Kim remains free—as was the dualist—to invoke his prior theoretical commitments in order to defend an account whereby qualia are nonphysical, but are spatially located. If Kim is justified in believing that qualia are irreducible, and he is justified in believing that qualia can be caused, and if he is persuaded by the pairing problem of the essentiality of spatial location for causal relations, then Kim can reasonably conclude that qualia must have a spatial location.[93]

It seems, then, that Kim's functional reduction is on par with substance dualism with respect to the pairing problem. If Kim rejects the dualist's attempt to locate minds in space on the grounds that such a move could not be adequately motivated, then he must also resist the move to locate qualia in space. If, on the other hand, he deems the location of qualia necessary in light of the pairing problem, then he must concede the dualist's justification in locating minds as well. The two moves are precisely analogous.

[93] Colin McGinn makes a similar move with respect to colors in "Another Look at Color," *Journal of Philosophy* Vol.93, No.11 (Nov. 1996) pp 537-553 Leopold Stubenberg alerted me to this similarity.

§4.2 Objections and Responses

Before concluding, I would like briefly to consider to a few objections that seem likely to arise in response to this parity claim. First, the physicalist might suggest that parity fails to hold because Kim is free to *eliminate* nonphysical qualia while the dualist, if she is to remain a substance dualist, cannot eliminate nonphysical minds. If this is true, then Kim can likewise eliminate any parity between his own account and that of the dualist with respect to the pairing problem. By eliminating qualia from his ontology, Kim can avoid the qualia pairing problem without having to postulate spatially located, nonphysical entities.

Of course, if qualia were so easily eliminated, it seems likely that Kim would have done away with them long ago. They are, after all, the only thing standing between his own "near enough" approximation of physicalism and physicalism *proper*. Qualia are not something that Kim embraces, but rather something that he reluctantly concedes. Accordingly, it is not at all clear to me that Kim is free to simply eliminate qualia without thereby losing some central feature of his account.

Alternatively, the physicalist might claim that qualia, unlike immaterial minds, are the kinds of things that *should* be located. As such, the location of qualia does not require as extensive a defense as would the location of immaterial minds. Qualia are, after all, aspects of human experience. Humans are located, and qualia *seem* to be located as well.

There are, in light of this, three ways in which qualia might plausibly be located.[94] First, we might locate a quale wherever the cause of that particular qualitative experience is located. If I am kicked in the shin and feel pain, I could locate that pain exactly where the bruise is located. This response will not do as a response to the pairing problem, however, for it presupposes the causation of qualia.

Second, we might speak of the "phenomenal location" of pain. If I have pain in my lower back, it has a "lower back" sort of feeling to it. However, this too seems unlikely to distinguish physicalism from dualism with respect to the pairing problem. Phenomenal locations are not spatial locations; they do not comprise any part of the space-time framework. Instead, the phenomenal location of a quale can tell us only what a pain *feels* like, where it *seems* to be. (Note that, if this were a sufficient response, the dualist could similarly appeal to the fact that one's mind seems to be located in the general area of one's head.) Unless phenomenal locations can be shown to be capable of discerning among qualitatively indiscernible relata, they cannot ground pairing relations.[95]

Finally, we might locate a quale "where its neural correlate is located."[96] If qualia can be correlated with neural states, then we can locate qualia according to these correlations. This would surely give us a physical location, and one that could suffice for causal pairing. Once again, though, this move is precisely analogous to the one that the

[94] These were suggested to me by Kim through an email correspondence pertaining to the qualia pairing problem. At the time he did not find any of the three to be a promising solution, though he might of course change his mind upon further reflection.

[95] In this respect, the decision to locate a quale in some phenomenal location is analogous to the decision to locate a mind in a purely intentional framework.

[96] Kim correspondence.

dualist ought to make—the location of the immaterial mind where the brain to which it is related is located. After all, substance dualists believe that minds belong to a unique body, and many believe further that mental changes correspond regularly with physical changes.

The dualist who locates the mind, roughly, where the brain is, is claiming that the mind is (a) nonphysical, (b) spatially coincident with a brain, but (c) not identical with the brain. Likewise, were Kim to affirm the location of qualia where their correlates are located, he would be claiming that qualia are (a) nonphysical, (b) spatially coincident with certain neural configurations, but (c) not identical with these neural configurations. The location of qualia where their (postulated) neural correlates are located will not do to distinguish Kim's position from that of the dualist's. Once again, parity holds.

There might be one way for Kim to respond which would not be analogous to any of the moves available to the substance dualist. An immaterial mind is purported to be a substance; a quale need not be a substantial entity. If Kim takes qualia to be properties, then he might assign a location to a quale according to whatever is instantiating that qualitative property.[97] What follows for the location of qualia if they are judged to be properties?

[97] There are reasons both in favor of and against the conclusion that Kim's irreducible qualia should be understood as properties. On the one hand, in Chapter One of *Physicalism, or Something Near Enough* Kim writes that "phenomenal mental properties are not functionally definable and hence functionally irreducible." (29) There, he clearly takes irreducible qualia to be mental properties. On the other hand, if qualia are properties, then property dualism is true; yet Kim maintains the falsity of property dualism in the conclusion of *Physicalism, or Something Near Enough*. (See, for example, pp158-159) He rejects property dualism because he believes it cannot account for mental causation, so the existence of *epiphenomenal* mental properties might be something he deems acceptable. Still, it is worth noting that the location of qualia in terms of property instantiation was not one of the options that Kim suggested the physicalist take in our correspondence over the qualia pairing problem, and Chapter Six of *Physicalism, or Something Near Enough* makes no reference to qualia as properties. It seems reasonable to conclude that Kim himself is as of yet uncommitted to any particular account of irreducible qualia. He states explicitly in the introduction to *Physicalism, or Something Near Enough* that the chapters were originally written as stand-alone essays, and that he has essentially preserved the independent nature of each essay.(pp. xi-xii) As such, we should not conclude any inconsistency in Kim's affirmation of irreducible mental properties in Chapter One and his rejection of property dualism in Chapter 6. Instead, as I suggest, a far more charitable interpretation is that Kim has not yet taken a firm stance as to how we should understand irreducible, epiphenomenal qualia.

First, if qualia are instantiated by physical objects, then they do not "stay outside of the physical domain;" they are instantiated *in* the physical domain, by physical objects. Still, by being instantiated by physical objects, it seems likely that qualia could be said to inherit the spatial location of those objects. If so, then this "location inheritance" could provide Kim with a novel solution to the qualia pairing problem—one to which the dualist cannot appeal.

If this position is to be defended, then there remains much work to be done. There are, it seems, two ways in which qualitative properties might be instantiated by a purely physical object. They might *supervene* on physical properties instantiated by that object, or they might be instantiated *directly*. I will begin with the first possibility. In light of the possibility of a qualia spectrum inversion, we have concluded that qualia do not strongly supervene on physical properties. As Kim writes, "There is a possible world that is like this world in all respects except for the fact that in our world, qualia are distributed differently." (170) However, the possibility of an inverted spectrum does not preclude weak supervenience from holding between qualia and physical properties. Perhaps, in the *actual* world, qualia supervene on physical properties.

As we saw in Chapter One, if Kim wishes to affirm this position, he needs to provide some explanation of why we should believe weak supervenience to hold despite the failure of strong supervenience. If we allow for worlds in which pain supervenes on C-fibers firing, for example, and we allow for worlds in which pain *fails* to supervene on C-fibers firing, what justifies our refusal to allow for a world in which pain sometimes does, and sometimes does not, supervene on C-fibers firing? As Simon Blackburn writes in "Supervenience Revisited",

> Why should the possible worlds partition into only the two kinds, and not into the three kinds? It seems on the face of it to offend against a principle of plenitude with respect to possibilities, namely that we should allow any which we are not constrained to disallow.[98]

Accordingly, if Kim is to affirm the weak supervenience of qualia on physical properties while denying strong supervenience, he should provide a defense of this claim. Absent some defense, it is difficult to see why we should conclude that, in the actual world, qualia weakly supervene on physical states. If he can provide a defense of this position, then perhaps his account will have an advantage over dualism with respect to the pairing problem.[99]

Suppose instead that qualia do not supervene on physical properties at all, but are instantiated directly. This, too, is a position that would require a defense. To see why, note that there are at least *prima facie* reasons to believe that physical objects cannot instantiate qualia. Perhaps the best articulation of the tension that arises given qualia instantiation can be found in Sellars's famous "grain argument."[100] Qualia are "ultimately homogenous"; we experience qualia as an indivisible unity. To use Sellars's example, consider the qualitative experience of observing a pink ice cube. Sellars writes,

> The manifest ice cube presents itself to us as something which is pink through and through, as a pink continuum, all of the regions of which, however small, are pink. It presents itself to us as *ultimately homogeneous*.[101]

---

[98] This is a simplified restatement of Blackburn's argument against the "ban on mixed worlds." See Simon Blackburn "Supervenience Revisited"(1985) reprinted in *Essays in Quasi-Realism*. (Oxford: Oxford University Press, 1993)

[99] Kim could, of course, justify his belief in weak supervenience in a way that is analogous to the dualist's decision to locate minds in space. This would not, however, help with the parity problem.

[100] I owe this reference to Leopold Stubenberg. Sellars himself believed this difficulty could be addressed on a physicalist picture. However, this does not preclude the dualist from appealing to the example as an indication of a perceived tension in the physicalist's picture.

[101] Wilfred Sellars *Science, Perception and Reality*. (London: Routledge and Kegan Paul, 1963) (Cited in Clark, Austen "The Particulate Instantiation of Homogeneous Pink" Synthese, 80:2 (1989:aug.) p.277)

The material of which an ice cube is comprised, however, is not so homogeneous. Further, *none* of the basic particles found in an ice cube are themselves pink. If heterogeneous matter instantiates homogeneous qualia, where and how does it do so?

It will not do to appeal to some physical arrangement of properties in order to locate the instantiation of qualia. We are supposing qualia not to supervene on physical properties, but to be instantiated directly. If some arrangement of physical properties could guarantee the presence of a quale in a physical object, then that quale would supervene on that arrangement of properties. It must be, then, that qualia instantiations—if they occur—are epistemically inaccessible to us. Note that this is *not* the case for functionalized mental properties, the physical instantiation of which we *can* detect. Provided we have identified the causal role specified by a property, we need only look for some physical realizer of that causal role in order to justifiably assert the instantiation of the functionalized mental property. No such process exists, or *could* exist, for qualia. As irreducible, non-supervening, epiphenomenal properties, they are not the sort of thing that we could observe from a third person perspective.[102]

If qualia are irreducible to physical properties, and if they fail to supervene on physical properties, then it follows that there is *no physical way for an object to be* that could constitute its instantiation of a given quale. (This is, of course, no surprise; Kim himself repeatedly notes the irreducibility—functional or otherwise—of qualia.) This does not render qualia instantiation *impossible*, but it certainly makes it difficult to ascertain how we could ever be justified in claiming to *know* that qualia are instantiated in a physical object. Difficult, that is, unless the physicalist appeals to his physicalist

[102] We could of course observe qualia from a first-person perspective. I will discuss this response momentarily.

commitments, his first-person experience with qualia, and the difficulties resulting from the pairing problem to justify this claim; this remains a possibility.

However, the matter with which we are presently concerned is not whether or not Kim can provide a solution to the qualia pairing problem, but whether or not he could do so in a way that distinguishes his response from that of the substance dualist. This response is precisely analogous to the dualist's decision to locate a mind on the basis of her dualistic commitments, her first-person experience with mental causation, and the difficulties arising from the pairing problem. As such, it will not do as a distinguishing factor between Kim's functional reductive physicalism and substance dualism with respect to the pairing problem.

Conclusion

Kim's formulation of the pairing problem poses a challenge to substance dualism, but it is a challenge to which the dualist can respond. Furthermore, the pairing problem also raises difficulties for the possibility of qualia that are truly epiphenomenal, rather than wholly causally irrelevant. It seems likely that Kim can adequately defend a response to this qualia pairing problem, but that, in doing so, he must respond in a way that is analogous to the response offered by the substance dualist. Just as the dualist ought to affirm the spatial location of immaterial minds in order to address the pairing problem, so too should Kim affirm the spatial location of nonphysical qualia. Neither response would be question begging and both, I suggest, are on equal footing with respect to having adequate theoretical motivation.

CHAPTER THREE:

SUBSTANCE DUALISM AND THE CAUSAL EXCLUSION ARGUMENT

In Chapter Two, we saw how an interactionist substance dualist can respond to one problem of mental causation: the pairing problem. This is good news, as some have supposed that the pairing problem renders untenable a substance dualist account of mental causation.[103] Unfortunately, the problem of mental causation for dualism does not end there. Instead, I am inclined to think that the *bigger* problem of mental causation for the substance dualist is the problem faced by nonreductive physicalists as well—the causal exclusion problem. According to the exclusion argument, there is simply no room for irreducibly mental causation. The causal closure of the physical world precludes the mental—insofar as it is nonphysical—from having any genuine, nonredundant causal efficacy in the world.

In the next two chapters, I will consider the ramifications of the causal exclusion argument for substance dualism. In the present chapter, I will show why the exclusion

[103] See, for example, *Physicalism, or Something Near Enough* p.92 (Princeton: Princeton University Press, 2005).

argument, typically directed against nonreductive physicalists, applies with equal force to substance dualists. I will then examine three responses to the argument, all of which proceed without disputing the causal closure of the physical world. In chapter four, I will turn my attention to the question of closure.

§1.1 Background Considerations

There have been quite a few formulations of the causal exclusion argument—sometimes called the "causal closure argument"—in recent years.[104] Though they differ in the details, the essential point remains: the causal closure of the physical world plus the plausible claim that mental causation is not systematically overdetermined render a nonreductive account of mental causation untenable. Unless mental causes are ultimately physical, they cannot act in the physical world.[105] In what follows, I will consider this argument in detail, noting the ways in which an interactionist substance dualist might respond.

In examining this argument, I will draw, in part, from the second chapter of Jaegwon Kim's *Physicalism, or Something Near Enough:* "The Supervenience Argument

[104] Some examples include: Alyssa Ney, "Can an Appeal to Constitution Solve the Exclusion Problem" *Pacific Philosophical Quarterly*, 88(4), 486-506. December 2007; Terry Horgan, "Mental Causation and the Agent Exclusion Problem" *Erkenntnis: An International Journal of Analytic Philosophy*, 67(2), 183-200. September 2007; E.J. Lowe, "Physical Causal Closure and the Invisibility of Mental Causation" and Andrew Melnky, "Some Evidence for Physicalism," both in Sven Walter and Heinz-Dieter Heckmann, eds. *Physicalism and Mental Causation*. (Exeter: Imprint Academic, 2003)

[105] In *Physicalism, or Something Near Enough*, Jaegwon Kim writes that "Unless we bring the supposed mental causes fully into the physical world, there is no hope of vindicating their status as causes, … the reality of mental causation requires reduction of mentality to physical processes, or of minds to brains." (Kim, 2005) p156

Motivated, Clarified, and Defended."[106] There, Kim offers a formulation that makes clear precisely from where the argumentative force of the causal exclusion argument stems. More specifically, Kim's "Supervenience Argument" centers upon the following two principles:

> *Closure*: If a physical event has a [sufficient] cause that occurs at *t*, it has a physical [sufficient] cause that occurs at *t*. (43)[107]
>
> *Exclusion*. No single event can have more than one sufficient cause occurring at any given time—unless it is a case of genuine overdetermination. (42)

What is particularly noteworthy about Kim's presentation is the appeal to *exclusion*. There, by explicitly stating a sufficient condition for overdetermination, Kim disambiguates two of the ways in which an opponent of the exclusion argument might respond.

More specifically, on a less precise formulation of the causal exclusion argument—for example, the rough and ready account I gave in the previous paragraph—it looks as if there are only two possible lines of response: reject the causal closure of the physical world, or accept the systematic overdetermination of mentally caused events. In light of Kim's formulation, however, we can see that there is a third option. A dualist who affirms *closure* need not embrace systematic overdetermination. She *might* do so, but she might *instead* reject Kim's exclusion principle, arguing that an event can have both a mental and a physical cause at the same time without thereby being overdetermined. We

[106] Kim 2005, pp32-69

[107] Kim's statement of *closure* does not include the word "sufficient," but—unless I am mistaken—sufficiency is implicit. If it is not, then the truth of both *closure* and *exclusion* is consistent with there being a physical event that has a sufficient mental cause and, simultaneously, only a partial physical cause. I don't believe that this is what Kim intends. Rather, by "cause," I take Kim to mean "sufficient cause."

will consider responses of both varieties.[108] (As for the rejection of *closure*, we will consider responses of that variety in chapter four.)

Of course, The Supervenience Argument is not specifically directed against substance dualists. Instead, as the name suggests, it takes as its target nonreductive physicalists who posit the supervenience of the mental on the physical. For that reason, it will be necessary to reformulate the Supervenience Argument so that it applies to the substance dualist. In doing so, I will employ Kim's causal exclusion principle, as well as his definition of the causal closure of the physical—a definition that is widely affirmed.[109] What follows is a causal exclusion argument inspired by, but certainly not identical to, Kim's Supervenience Argument.

§1. 2 The Causal Exclusion Argument Against Substance Dualism

There are a number of ways one might formulate a causal exclusion argument against interactionist substance dualism.[110] In his Supervenience Argument, Kim begins

[108] There is a third line of response as well: one might distinguish *bad* (or *redundant*) overdetermination from *acceptable* (or *nonredundant*) overdetermination. She could then concede that there is no systematic *bad* overdetermination, but deny that the kind of overdetermination that is involved with mental causation is problematic. (For a discussion on this, see Alyssa Ney, "Can an Appeal to Constitution Solve the Exclusion Problem" *Pacific Philosophical Quarterly*, 88(4), 486-506. December 2007, especially pp.487-488.**)** This position is genuinely distinct from that which denies *exclusion*—the two differ with respect to how they conceive of overdetermination, for example—but, as a response to the causal exclusion argument, they amount to the same thing. Whether we call it an acceptable form of overdetermination or deny that it is overdetermination at all, the point remains: whatever it is that mental causes do with respect to physical causes, there is no reason to believe that they cannot do so in a widespread, systematic fashion.

[109] In this chapter, Kim's *closure* is the only closure principle that I will consider in detail. I do this, primarily, because *closure* is a good representative of most closure principles. That is, it is fairly typical and carefully formulated. I will, however, briefly consider alternative ways of understanding causal closure in §4.2 of this chapter and in Chapter Four.

[110] Kim's Supervenience Argument against nonreductive physicalism is given three formulations. Similarly, we might formulate the argument against substance dualism in a variety of ways—depending, for

with the assumption that one mental event is the cause of another mental event. In contrast, I will begin with the assumption that a mental event is the cause of a *physical* event.[111] For this reason, if the argument succeeds, it does so only against mental-to-physical causation; it says nothing about the possibility of mental-to-mental causation. However, while mental-to-mental causation might satisfy *some* dualistic accounts of mental causation—for example, parallelism—it will not, of course, suffice for interactionism. An interactionist substance dualist is committed to the possibility of mental-to-physical causation.

In light of this, I offer the following statement of the causal exclusion argument against interactionist substance dualism:

(1) Suppose mental event M causes physical event P at *t*. (*for reductio*)
(2) P has a sufficient physical cause at *t* as well, call it P*. (*closure*)
(3) M is not identical with P*. (substance dualism)
(4) P cannot have more than one sufficient cause at *t*—unless this is a case of genuine overdetermination. (*exclusion*)
(5) This is not a case of genuine overdetermination.
(6) Then either P* or M is the cause of P, but not both. ((1)-(5))
(7) P*, not M, is the cause of P.
(8) If M causes P at *t*, then M does not cause P at *t*. ⊗

The formulation differs from that of Kim's Supervenience Argument, but the message is the same: an irreducibly mental cause has no place in the physical world.

Before considering specific responses to the causal exclusion argument, it will be helpful to look at the argument in a bit more detail. Premises (1), (3), and (6) are the easiest to defend. Premise (1) is nearly nonnegotiable; if interactionist substance dualism

---

example, upon whether we begin with mental-to-mental causation or mental-to-physical, or on how we conceive of the causal closure of the physical world.

[111] In general, I will assume that causation is a relation that obtains between events. This assumption will, however, be reconsidered in section 4.

is true, then, presumably, something like this happens from time to time.[112] Assuming that causation is a relation that obtains between events, and that the mental is causally efficacious in the physical world, a mental event must sometimes be the cause of a physical event. Premise (3) is similarly established. If substance dualism is true, then a mental event—which will involve a mental substance—is not identical to *any* physical event. It is, therefore, not identical to *this* physical event. Premise (6) follows from (1)-(5).

What about (7)? It is worth noting that, technically, (7) need not follow from (1)-(6). Instead, we might claim that M *and not P** is the true cause of P. This will not do, however, for the following reason: If we reject the claim that P* is the cause of P, *closure* demands that we find some *other* physical cause of P at *t*. Call this P2. By *exclusion*, either M or P2 must be rejected as the cause of P at *t*. Once again, we *could* reject P2, but it should by now be clear that this response is a nonstarter. If (1)-(6) are affirmed, (7)—or something quite like it—must *eventually* follow.[113]

That leaves (2), (4), and (5). Here we find the heart of the argument: a causal closure principle (2), a causal exclusion principle that sets the parameters for overdetermination (4), and the claim that a mentally caused event will not, *in general*, be overdetermined (5). If a dualist is to succeed in responding to the causal exclusion argument, she will likely do so by rejecting one of these three premises.

The question, of course, is which premise to reject. If the dualist rejects (2), she denies a widely held causal closure principle. This might amount to a rejection of the

[112] I say "nearly" nonnegotiable because, as it turns out, at least one interactionist substance dualist might reject this premise. As we shall see in section 4, E.J. Lowe offers an account whereby all *event* causation is physical causation, while mental causation is *fact* causation.

[113] Kim notes this as well. (Kim 2005, p43)

causal closure of the physical world, or it might only consist of a rejection of *this* closure principle. In either case, a dualist who rejects (2) will have her work cut-out for her. According to many highly regarded philosophers, a rejection of *closure*—or some similar principle—is tantamount to the rejection of "science," and is, therefore, not really a tenable alternative. (Others are, of course, not so sure.) A substance dualist who wishes to preserve the appearance of scientific respectability will, therefore, be better served by a rejection of either (4) or (5). In the next chapter, we will revisit the question of the compatibility of causal closure with interactionist substance dualism. The remainder of this chapter, however, will focus on responses that accept (2) and reject, instead, either (4) or (5).

What is involved in rejecting either of these premises? In order to reject (4), a dualist must maintain that the conditions laid-out in *exclusion* do not suffice for overdetermination. Instead, it must be possible for an event to have more than one complete cause at a given time *without* thereby being overdetermined. A rejection of (5), on the other hand, need not involve a rejection of *exclusion*. One might agree with Kim's conception of overdetermination, but disagree with the implicit claim that mentally caused events are not systematically overdetermined. Without this implicit premise, there is simply no reason to conclude that (5) is true. If such widespread overdetermination is compatible with the causal efficacy of the mental, then M may very well be overdetermined.

If either (4) or (5) can plausibly be rejected, then neither M nor P need be excluded, (and so (6) does not follow,) and the argument fails. It is this approach, the rejection of either *exclusion* or the claim that mental causes are not systematically

overdetermined, upon which the remainder of this chapter will focus.[114] (We will, however, briefly consider a third alternative—the rejection of mental event causation in favor of mental *fact* causation—in Section Four.) In the end, I hope to show that the substance dualist can respond to the causal exclusion argument while affirming the causal closure of the physical world. At the same time, I maintain that the concessions that a dualist must make in order to affirm *closure* take their toll on the resulting dualistic account, ultimately undermining much of what is often held to be desirable about dualism. I conclude by suggesting that the dualist consider rejecting the causal closure of the physical after all, despite the fact that the causal exclusion argument fails to necessitate such a rejection.

§2. Embracing Overdetermination: Eugene Mills

In "Interaction and Overdetermination," Eugene Mills defends substance dualism from the causal exclusion argument by embracing systematic overdetermination.[115] According to Mills, all mental causes are overdetermining causes, but an overdetermining mental cause is a cause nonetheless. He offers the following illustration:

> Suppose my believing that Bill Monroe is the father of bluegrass music causes me to raise my arm. Someone asks from a concert stage, "How many of you believe that Bill Monroe is the father of bluegrass?" and I raise my arm in response. Call the proposition that Bill Monroe is the father of bluegrass B. My believing B is a mental event and, ex hypothesi, is causally sufficient in the circumstances for the (physical) arm-rising. Given this assumption, physical closure tells us that the arm-rising also has a physical sufficient cause. Hence the proper conclusion, given

[114] I refer the reader to footnote 7 for an alternative way of construing a dualistic response to this argument.

[115] *American Philosophical Quarterly* 33.n1 (Jan 1996): pp105(13).

> our assumption of belief-dualism, is that the arm-rising is causally overdetermined by physical and mental events. (106)

Mills's reasoning is fairly straightforward, and parallels that of the physicalist running the exclusion argument. If we have good reason to think that beliefs are not physical entities, and we have good reason to think that our beliefs sometimes cause us to act, then, in light of causal closure, we should conclude that these belief-causes are overdetermining ones. (Unlike the physicalist running the exclusion argument, Mills does *not* find overdetermination problematic, and so rejects the push towards reduction.)

Mills begins by noting that his position is compatible with *exclusion* and *closure*; both technically allow for overdetermining mental causes. However, merely noting the theoretical possibility of overdetermination will not suffice as a *defense* of an overdeterministic position, for systematic overdetermination, he concedes, "is widely thought wildly implausible." (105) For this reason, he begins with a defense of the claim that mental causes are overdetermining ones, and then proceeds to respond to a series of objections against such a position.

To show why a mental cause ought to be understood as an overdetermining one, Mills offers the following four counterfactual statements. (In what follows, B is the belief that Bill Monroe is the father of bluegrass music, and P is the physical sufficient cause of the arm-rising.)

> S_1 If I hadn't believed B, then if I had, the arm-rising would have occurred.
> $S_{2\,If}$ P hadn't occurred, then if it had, the arm-rising would have occurred.
> O_1 If P hadn't occurred but my belief had, the arm-rising would have occurred
> O_2 If my belief hadn't occurred but P had, the arm-rising would have occurred.

If all four of these counterfactuals are nonvacuously true, then, according to Mills, the arm-rising should be deemed overdetermined. It is, however, important that their truth not

be vacuous. After all, if the belief B just *is* the physical cause P, then O_1 and O_2 will be trivially true. It will not, of course, follow that the arm-rising is overdetermined; no event can be overdetermined by a single cause. Henceforth, by "true" I will mean "nonvacuously true."

Consider first S_1: "If I hadn't believed B, then if I had, the arm-rising would have occurred." If S_1 is true, then B is likely a cause of the arm-rising. To see why, it will be helpful to note the inadequacy of a closely related counterfactual, S_0: "If I had believed B, the arm-rising would have occurred" (3) At first glance, S_0 seems also to support the efficacy of B in bringing about the arm-rising, but this is not the case. S_0 is too easily confirmed. If we begin with the assumption that my belief, B, preceded the arm-rising, then we begin with the assumption that both B and the arm-rising occurred in the actual world. The actual world being the nearest possible world to itself, S_0 is therefore true, but its truth tells us nothing about the relationship between B and the arm-rising. Suppose I was wearing a yellow shirt at the time of the arm-rising. It follows that, if I had been wearing a yellow shirt, the arm-rising would have occurred; it does *not* follow that my wearing the yellow shirt was a *cause* of the arm-rising. We cannot, therefore, look to S_0 to support the causal efficacy of B.[116]

S_1, on the other hand, is not so easily confirmed; to determine its truth value, we must consider the closest world in which I did *not* have the belief in question.[117] If the closest B world to *that* world is one in which the arm-rising occurs, then we have good reason to believe that instances of B (in the right circumstances) cause arm-risings. We

[116] One might take this as evidence of the fact that we ought to revisit our counterfactual semantics. As Alvin Plantinga pointed-out to me, "'If China were a large country, I'd be typing now' doesn't seem to be much of a counterfactual."

[117] There need not be any *one* closest world fitting this description. Here, and throughout this paper, by "closest world" I will mean "closest world or set of worlds."

therefore have reason to believe that, in the actual world, B is a cause of the arm-rising. Likewise, S_2 –if true—gives us reason to believe that P is a cause of the arm-rising.[118]

O_1 and O_2 ensure that B and P are *independent* causes of the arm-rising, rather than two parts of a single cause. Once again, it is crucial that these counterfactuals are genuinely counter to the assumed facts, and are therefore not vacuous. If, in the world closest to our own in which B fails to obtain, P nevertheless causes the arm-rising, we can be confident that P is not dependent upon B for its causal efficacy. Likewise, if B would have caused the arm-rising in P's absence, then B's causal efficacy is independent of that of P.

We are now in a position to see how the collective truth of these four counterfactuals would indicate that the arm-rising is overdetermined. From S_1 and S_2 we learn that the arm-rising has two causes. From O_1 and O_2 we learn that these causes operate independently of one another. While Mills does not claim that B and P occur simultaneously, he certainly could do so, and likely *would* if that is what *closure* requires. (After all, he is attempting to show what follows from the conjunction of belief dualism, mental causation, and physical causal closure.) If Mills's four counterfactuals are all true, then we have good reason to believe that the arm-rising is, in fact, causally overdetermined. Furthermore, on the plausible assumption that the arm-rising represents a typical case of mental causation, it follows that we have good reason to believe in the widespread, systematic overdetermination of mentally caused events.

[118] It is, perhaps, better to say that if the counterfactual is true, then whatever evidence counts in favor of the truth of the counterfactual counts in favor of the causal efficacy of P (or B). As we will see, however, even this is not necessarily the case. It might turn out that *some* reasons for believing S1 or S2 have little bearing on the plausibility of the claim that B, or P, is a cause of the arm-rising. We will return to this subject shortly.

The question, then, is whether or not one or more of the counterfactuals can be shown to be false. As Mills notes, "no one will worry about the causal influence of P, so the real targets are S_2 and O_2." (107) In order to defend the causal influence of B, however, Mills notes that we ought to examine the *reasons* for which S_1 and O_1 are so readily accepted. The primary reason, according to Mills, can be understood as follows: when evaluating the proximity of possible worlds, we hold the laws of nature constant whenever possible. If, as we have postulated, P is a sufficient cause of the arm-rising, it is so in accordance with of the actual laws of nature. Furthermore, the nearest possible world in which P fails to obtain will not be very far away at all; surely in this world the laws of nature are the same as those of the actual world. But in moving to the nearest world to *that* world in which P *does* obtain, we likewise can and should hold the laws of nature constant. We may therefore conclude that P will suffice for the arm-rising, and that S_2 is true.

The truth of O_2 is a bit more complicated, but not by much. Suppose that P occurs and B does not. If, as Mills deems likely, there is a psycho-physical law linking occurrences of P with occurrences of B, then in (P & ~B) worlds, this law fails to obtain. Still, in evaluating these worlds, it remains the case that the ones with fewer nomic transgressions will be closer than those whose laws vary dramatically from our own. Thus, in the *nearer* (P&~B) worlds, the physical laws will remain constant despite the violation of the aforementioned psycho-physical law. P will therefore cause the arm-rising in accordance with the relevant physical laws, and O_2, can be seen to be true.

Furthermore, according to Mills, if we accept S_2 and O_2 on these grounds, we ought to accept S_1 and O_1 for similar reasons. He writes the following:

> Consider S_1. If w is among the nearest worlds to ours in which my belief does not occur, it has the same laws of nature as the actual world--including any psycho-physical laws--and the nearest worlds to w in which my belief does occur have these laws as well. But the nearest worlds to w in which my belief occurs and in which actual laws of nature hold are, intuitively, worlds in which the arm-rising also occurs. For worlds in which my belief is accompanied by some physical event that causes the arm-rising preserve actual laws, whereas worlds in which my belief is unaccompanied by any such physical event do not. (107)

Suppose that B and P are correlated via a psycho-physical law of nature. Then the nearest ~B world will also be a ~P world, for to suppose otherwise would be to suppose that the psycho-physical law of nature correlating the two fails in that world. (We *need* not assume this nomic transgression, and so we *ought* not do so.) By the same reasoning, the nearest B world to *that* world will be a P world, and P will, in accordance with the laws of nature, cause the arm-rising. Thus, if we accept the importance of the laws of nature in evaluating the proximity of possible worlds, (and if we accept the existence of psycho-physical laws of nature), we must conclude that S_1 is true.

Once again, O_2 is slightly more complicated, but not by much. In the nearest (B & ~P) world, some psycho-physical law is violated, and so the laws of nature are not exactly as they are in the actual world. Still, the *nearest* world in which this occurs will be one that is otherwise quite similar to ours; for this reason, we may assume that it will be a world in which the arm-rising occurs, as it does in the actual world. At the same time, it will be a world in which *closure* holds, and so the arm-rising will have *some* physical cause. Mills writes:

> Intuition insists on both of the following claims: (1) even if my body hadn't been in its actual physical state, my arm still would have risen had I believed B; and (2) my arm would have risen only if my body were in some physical state causally sufficient for its rising. (109)

In this way, Mills concludes, S_2 and O_2 can be justified in much the same way as S_1 and O_1.

§2.2 Objections To Mills's Overdeterminism

If Mills is correct about both the significance and the truth of these four counterfactuals, then the committed dualist ought to conclude that mentally caused events are systematically overdetermined. That said, in order for the counterfactuals to do the work for which they are intended, it matters not only *that* they are true, but also *why* they are true. To see why this is the case, consider S_2: "If I hadn't believed B then, if I had, the arm-rising would have occurred." At first glance, this counterfactual seems to tell us that B is a cause of the arm-rising. After all, as was previously stated, it looks as if there is some counterfactual dependency between B and the arm-rising. Despite this initial appearance, however, Mills's defense of the truth of S_2 presents a somewhat different picture. We are encouraged to believe that S_2 is true because the nearest B world (*w*) to the ~B world (*w'*) will be one in which P causes the arm-rising.[119] It follows that an arm-rising occurs in *w*, and the counterfactual is therefore true.

The trouble is, this tells us very little—indeed nothing—about *B's* causal efficacy. In "Why The Exclusion Problem Seems Intractable, and How, Just Maybe, To Tract It,"[120] Karen Bennett considers this line of reasoning. She notes the popularity of this response among "causal compatibilists"—those who would defend the compatibility of purportedly competing mental and physical causes. She writes: "Now, this had better not

[119] Mills does not explicitly state that it is P, rather than something like P, that causes the arm-rising in the world being considered. However, if he is correct about preserving psycho-physical laws, than P ought to be there to do the causal work. In any case, the point remains even if it is another physical state, P', that plays this role instead.

[120] Nous 37:3 (2003) 471-497

be the compatibilist's only reason for thinking that [the relevant counterfactual] is true. It is compatible with the mental event or property being utterly epiphenomenal." (482) To see why Bennett is correct, suppose that, in the actual world, parallelism is true. Then mental states and physical states correlate perfectly with one another, but there are no mental causes of physical events. If, as Mills suggests, we hold the laws of nature constant when evaluating the nearness of worlds, then the nearest world in which my belief does not occur (*w'*) will be a world in which the arm-rising nevertheless occurs—caused, in accordance with these laws, by P. The nearest B world to *w'*, (*w*) will likewise contain the arm-rising. It follows, then, that S_2 is true. What does *not* follow is the fact that B is a cause of the arm-rising. After all, we are assuming that the actual world is one in which parallelism is true.[121] Given the truth of parallelism, we know that B is not the cause of any physical event. It is not, therefore, the cause of the arm-rising.[122]

The mere *correlation* of B and P cannot suffice to ensure the causal efficacy of B, no matter how reliable a cause *P* is. For this reason, although Mills presents a tenable defense of the *truth* of S_2, that defense undermines the *usefulness* of this counterfactual in guaranteeing B's causal efficacy. Furthermore, if we encounter these difficulties with S_2, then O_2 surely fares no better. There, you may recall, we are asked to believe in the independence of B's causal efficaciousness on the grounds that, should B occur in P's absence, *some other physical cause* would suffice for the arm-rising. At this point, it should be clear why this line of reasoning is inadequate.

Mills anticipates this objection, which he summarizes as follows:

---

[121] It seems worth noting that, an this hypothesis, both *w* and *w'* will be worlds in which parallelism is true as well. As Mills himself notes, we ought to hold the psycho-physical laws constant just as we do the physical ones.

[122] If causation just is constant conjunction, then this objection fails. Still, it would be quite a surprise if it turned out that interactionist dualism required one to be a Humean about causation.

> I argued above that had I believed B but P had not occurred, some other purely physical event would have occurred that would have been causally sufficient for the arm-rising. But this suggests that the mental event could not produce the arm-rising in the absence of physical causes. It seems odd, then, to call it a sufficient cause. (113)

His response to this objection is fairly brief. He writes, "This objection errs in moving from "would not" to "could not."(113) While Mills concedes that there are surely *some* physical worlds in which B succeeds in bringing about an arm-rising absent any physical sufficient cause, such worlds are "far enough from actuality that they have no bearing on the counterfactuals discussed above." (114) The actual world is one in which causal closure holds, and every arm-rising has a physical sufficient cause. Any world near enough to the actual world to be relevant to S_2 and O_2, then, must also be a world in which arm-risings have physical sufficient causes. For this reason, the worlds that we look to in order to confirm S_2 and O_2 are worlds in which P, or something like it, causes the arm-rising.

There is a sense in which this response simply misses the point of the objection. The claim is not that we would be better served by consulting a world that is very, very far away in order to defend these counterfactuals, but rather that, if such a world really *is* so far away as to be "irrelevant," then it looks like maybe B isn't a cause of the arm-rising after all. At the very least, even if B *is* a cause of the arm-rising, that is not something we learn from S_2 so conceived. For the truth of S_2 to be significant, we should be able to affirm it without knowing anything at all about the presence or absence of a physical cause of the arm-rising. The presence of B should itself suffice. If it doesn't, if instead we need to appeal to P (or something like it) as evidence, then S_2 will not be of much use to the dualist. Even if it is true, it will tell us nothing about *B's* efficaciousness.

Bracketing for a moment the aforementioned concerns, suppose that we grant the truth and the significance of S_2 and O_2. What follows? Well, both B and P seem to be causes of the arm-rising. If substance dualism is true, and B cannot be identified with P, then the arm-rising is overdetermined. However, as was previously noted, the implications of this claim extend far beyond a single event. If the arm-rising is a typical example of a mentally caused event, as Mill maintains, then *all* (or at least most) mentally caused events are overdetermined. Mills writes,

> I've argued against the oddity of psychophysical overdetermination. It might be responded that while occasional overdetermination needn't be odd, it does seem pretheoretically bizarre that there should be systematic, thoroughgoing causal overdetermination, which my view seems to require. (115)

Widespread, systematic overdetermination is, in the words of Jaegwon Kim, "at best extremely odd."[123] Many hold that if it can be avoided, through, say reduction, then it should be.

Mills offers two responses to this objection. First, he suggests that the oddness, insofar as there is any, is more of a problem for the physicalist than for the dualist:[124]

> What, exactly, is "systematic" about psycho-physical overdetermination? Not that every (caused) bodily motion is overdetermined by physical and mental causes: many physically caused bodily motions have no mental causes whatsoever. Nor is it true that mental events apt for causing bodily movement always do so, as paralysis shows. The point can only be that whenever a mental event causes a physical one, a physical cause operates as well. If there is oddity here, it is that such physical back-ups should always exist, not that mental events have causal influence on the physical. At least, this is the only oddity that I can see in the charge of "systematic overdetermination." And it impugns physical closure if it impugns anything. (115)

[123] Kim, Jaegwon "Mechanism, Purpose and Explanatory Exclusion" in *Supervenience and Mind*. Cambridge: Cambridge University Press. (p247)

[124] Mills's line of reasoning with respect to this question is unique, though I will claim that it is ultimately unsuccessful. I include the entirety of the following, admittedly lengthy quotation in order to ensure that I do not inadvertently misrepresent his position.

If Mills is correct, then the threat of systematic overdetermination is no threat to dualism at all. Instead, it is the causal closure of the physical that is in danger of being undermined.

What should we make of this line of reasoning? I can see at least two ways of understanding Mills here. (1) The dualist who affirms the causal efficacy of beliefs should reject *closure*, for the physical "back-up" that is supposed to accompany every mental cause would be extraneous, or (2) *closure* should be maintained, but mental causation should not be blamed for the apparent oddness of overdetermination. Instead, it should be seen as the (perhaps odd, but nevertheless true) result of the causal closure of the physical. In light of Mills's stated commitment to the truth of causal closure, (2) seems like the most likely reading of the above response. The oddness of overdetermination, if Mills is correct, stems from the presence of the overdetermining *physical* causes, not the mental ones.

That said, it's hard to see why we should believe this. Mills offers, as evidence, the following claims: (a) not every (caused) physical event has a mental cause, (b) not every potential mental cause is successful, and (c) every mental cause that *does* succeed in causing a physical event is accompanied by a physical cause. The trouble is, it's not at all clear that (a)-(c) support Mills' claim. Consider: if (a) is true, then physical causes can, and do, operate independently of mental ones. If (b) is true, then mental properties sometimes fail to bring-about their effects—and in the stated case of paralysis, they fail because there is no corresponding physical cause of the intended effect. If (c) is true, then for every mentally caused event there is one apparently redundant cause. Why, though, should we conclude from all of this that it is the *physical* cause that is the redundant—or

"odd"—one? What justifies the claim that the physical cause is rightly called a "back-up," rather than the true, (and perhaps only), cause of the ensuing event?

The principle of the causal closure of the physical states that every physical event that has a sufficient cause (at *t*) has a sufficient *physical* cause (at *t*). If we are to affirm *closure*, then, we cannot consistently claim that it is "odd" for there to be a physical cause of any physical event—even if we wish also to say that the event in question had a mental cause. For this reason, if the substance dualist accepts *closure*, then the burden of proof with respect to the oddness of overdetermination is not so easily shifted to the physicalist. Any oddness, insofar as there *is* oddness, sits squarely on the dualist's shoulders.

Mills does, however, have a second response to this objection: oddness, he notes, is not much of a charge. For all we know, the world *is* an odd sort of place. Without some additional reason to believe that widespread overdetermination is *impossible*, or even unlikely, there is little force to the "systematic overdetermination objection." Mills writes, "Oddity is hardly odd in science or the world. The question is whether overdeterministic interactionism is true." (117)

Furthermore, while one might appeal to simplicity to argue against the likelihood of systematic overdetermination, such an appeal holds little weight when lodged against a committed dualist. Simplicity is a useful criteria for theory choice *all things being equal*. As Mills notes, however, all things are decidedly *not* equal. The dualist is a dualist for a reason, and presumably her evidence for dualism must be weighed against any concerns about simplicity. A dualist who has good reason to believe in the existence of distinct mental causes should not, therefore, be swayed by concerns about the oddness of overdetermination, or the bulkiness of an overdeterministic account.

This, I think, is a reasonable response for a committed interactionist dualist who wishes also to affirm *closure*. Sure, it's a bit strange to think that every mental cause is overdetermined by a physical one, but if overdetermination is the only way to preserve dualism, psycho-physical causation and *closure*, then oddness alone should not suffice as a deterrent. In "Kim's Master Argument," Ted Warfield and Thomas Crisp offer the following response to Kim's exclusion argument:[125]

> Kim seems to presuppose that we shouldn't take seriously the possibility that every case of mental to physical causation involves a case of overdetermination. We suggest, though, that anyone committed to Closure and Property Dualism…should take this possibility quite seriously indeed. For Closure and Property Dualism together imply that every case of mental to physical causation is a case of causal overdetermination. (313)

If embracing overdetermination is the only way for the dualist—property or substance—to affirm both dualism and *closure*, then oddness alone should not deter her from doing so.

That said, the problem of justifying the causal efficaciousness of mental causes remains. As we have seen, the counterfactuals which were intended to support the claim that B, and not just P, is a cause of the arm-rising seem not to do so. If the dualist is to maintain that B is a *cause* of the arm-rising, and that beliefs in general are causally efficacious, she should be able to justify this in some way. This seems particularly important for the dualist who affirms *closure*. Having granted that the arm-rising has a physical sufficient cause, such a dualist should be prepared to give some reason for affirming the existence of a second, distinct cause. Note that evidence in favor of belief dualism—that is, in favor of the fact that beliefs are ontologically distinct from any physical state—will not do. Belief dualism is wholly compatible with belief

[125] NOUS 35:2 (2001) 304–316

epiphenomenalism.[126] For this reason, while Mills is correct that overdetermination might be both odd and true, he has not given sufficient reason to believe that it *is* true. Indeed, by emphasizing the irrelevance of worlds that contain instances of independent psycho-physical causation, he seems instead to indicate that *this* world, at least, is not one in which significant mental causation occurs.

Mills's position is technically a tenable one. There is nothing incoherent about supposing that mental causes are overdetermining ones, nor is there anything devastating about the oddness that such systematic overdetermination would entail. Still, there is a difference between a tenable position and a plausible one. For this to be a plausible position, it needs to be supplemented with some justification in favor of the reality of mental causation—justification that doesn't undermine either the ontological distinctness of the mental or the causal closure of the physical world. Absent such evidence, it seems that if the dualist can find another response, she ought to. In the words of E.J. Lowe, "I take it that most interactionist dualists would not wish to resort to this strategy if possible, as it looks suspiciously ad hoc."[127]

---

[126] Mills responds to a total of four objections, one of which may strike the reader as relevant here. The "Horganic Objection" offers the following problematic scenario for the dualist: Just as P is a sufficient cause of the arm-rising, there is some alternative physical cause, P', which would suffice for the arm's *not* rising. In the right circumstances, P' causes the arm to stay put, so to speak. According to this objection, the following is a true counterfactual: If (P' & B), then the arm-rising does not occur. But if this is true, then B seems clearly not to be a cause of the arm-rising, since it's occurrence is compatible with the arm-rising's nonoccurrence. Mills rejects the truth of this counterfactual. He notes that, in considering the proximity of worlds, we must take into account both physical laws and psycho-physical laws, and rejects the claim that the former always trump the latter. Instead, he suggests that B may very well cause the arm-rising—even given P'. If it were true that B would bring about the arm-rising even given the presence of P', then the dualist would be justified in affirming B's causal efficacy. The trouble, as I see it, is that Mills has given us little reason to believe that this could be the case. By defending the efficacy of B solely in terms of P-like physical states, Mills seems to undermine any credibility that this response might have had. For this reason, I have not emphasized this aspect of Mills's account.

[127] (Lowe 2000) p.572

§2.3 Final Reflections on Mills: An Analogy

An account of mental causation might be ad hoc, insufficiently motivated, odd, and—as Mills himself notes—it might nevertheless be *true*. It is worth asking, then, what would follow from the truth of Mills's account. If dualistic mental causation is real, but all mental causes are overdetermining ones, has the dualist won the day? Well, she has and she hasn't. Suppose a fervent believer in Santa Claus were to discover that there really is such a person as Santa Claus, but that all Christmas present are purchased, wrapped, and deposited under the Christmas tree by regular old parents. There might be *some* satisfaction in learning that Santa exists, and *some* sense in which the believer was right to defend his existence, but it hardly seems true to say that she has won the playground debate. Instead, it seems the skeptical children were *mostly* right, and she won—if at all—on a technicality. This, I suggest, is the position that the dualist would be in should Mills' account turn out to be the correct one.

To see why, consider the following analogy: Suppose there is a very strict, very powerful Sergeant, and every order issued by the Sergeant, *without exception*, is heeded by his subordinates. Indeed, his subordinates are so submissive to his will that not one of them ever performs an action without first having been ordered to do so by the Sergeant. Now suppose that, on occasion, the Sergeant brings his 4 year old son to work with him. Before doing so, he prepares the boy by telling him all of the orders that he will issue, and then prompts the boy to issue the same orders precisely when he does. Suppose further that he begins every day with a 5:30am order to stand at attention. When present, the boy of course issues this order as well. On those days, does the boy cause the subordinates to stand at attention?

Well, by Mills's reasoning, it seems that he does. To see why, consider the following counterfactual: "If the boy hadn't issued the 5:30 order to stand at attention, then, if he had, it would have been heeded."[128] Now consider the nearest possible world in which the boy does not issue this order to the troops. Is this a world in which *nobody* issues the order? Or in which a less-powerful Sergeant does so? Not likely. It seems, instead, that the nearest world in which the boy doesn't issue the order is a world in which the Sergeant issues the order anyway, just as he does every morning. Likewise, the nearest world to *that* in which the boy issues the order is hardly one in which he acts in the Sergeant's stead, but rather one in which the boy and the Sergeant act in unison—and so, of course, the order is heeded! If we follow Mills's reasoning, then, the boy does indeed cause the troops to stand at attention, and to march, and to do all manner of things.

There are, then, two questions: First, do we *really* want to call this causation? And second, even if we *do*, what is it worth, exactly? If we concede that the 4 year old causes the troops to fall in line, we don't thereby believe that the boy makes any kind of *difference* to how things go. It's not as if some of the soldiers would have *disobeyed* the sergeant but, thanks to the presence of the 4 year old, they obey.[129] What, then, do we gain by conceding that the boy's commands, though overdetermining, are nevertheless causally efficacious?

---

[128] By way of reminder, this analogy is based upon Mills's reasoning in the following passage: "Consider S1. If w is among the nearest worlds to ours in which my belief does not occur, it has the same laws of nature as the actual world--including any psycho-physical laws--and the nearest worlds to w in which my belief does occur have these laws as well. But the nearest worlds to w in which my belief occurs and in which actual laws of nature hold are, intuitively, worlds in which the arm-rising also occurs. For worlds in which my belief is accompanied by some physical event that causes the arm-rising preserve actual laws, whereas worlds in which my belief is unaccompanied by any such physical event do not.

[129] Indeed, we are assuming that no soldier ever disobeys the Sergeant. This is, after all, what closure tells us about the sufficiency of the physical causes that coincide with mental ones.

Whatever it is, it does not seem to be what the dualist was after when seeking to defend mental causation.[130] But this is precisely what Mills offers the dualist—mental causation that is wholly derivative of physical causation, and wholly impotent with respect to any kind of deviation from the path set-out by the physical causal order. Whether or not we choose to call this mental causation is, it seems to me, a question of semantics; the heart of the matter has already been decided. Insofar as Mills's account gives us mental causation, it is a paltry kind of causation at best, and hardly seems worth fighting for.

§3. Against *Exclusion*

Mills's response to the causal exclusion argument was to embrace systematic overdetermination while denying that such overdetermination is problematic. As we have seen, this is a theoretically tenable response to the exclusion problem, but absent supplementation it is not a particularly plausible (or attractive) one. Perhaps, then, the dualist would be better served by a rejection of *exclusion*. In what follows, we will consider a response of this variety.

It is worth noting that this is the strategy most often invoked by nonreductive physicalists in response to the causal exclusion argument. In *Why the Exclusion Problem Seems Intractable, and How, Just Maybe, To Tract It*, Karen Bennett describes the situation faced by most nonreductive physicalists when confronted with this argument: [131]

[130] (At the very least, it's not what *this* dualist is after.)

[131] *Nous* 37:3 (2003) 471-497; 472-473

> They want to hold fixed completeness, the distinctness of the mental and physical, and the causal efficacy of the mental, and *still* deny overdetermination. What they want to deny, then, is the claim that lurks in the background—that no effect can have more than one sufficient cause unless it is overdetermined.

This lurking claim is, of course, *exclusion*.[132]

Bennett goes on to note that, for the physicalist, the typical approach is to look for some kind of "tight relation" that obtains between the purportedly competing physical and mental causes. Often, the relation that is invoked is that of realization. The argument proceeds roughly as follows: "If the mental cause is *realized by* the physical cause, then the charge of overdetermination seems less plausible. If P *realizes* M, then P and M are distinct, but it would be odd to think of them as competing with one another for causal efficacy. Physical properties and the mental properties that they realize should not be seen as competitors. "[133] As I said, this is *roughly* how the argument goes. The difficulty, of course, is in the details. After all, a particularly tight relation between two causes—*identity*, say—could explain why they always coincide; the trick is to do so without collapsing into a reductive account.

For the substance dualist the danger is even greater, for she affirms an even more radical distinction between of mental and the physical. In "Physical Causal Closure and the Invisibility of Mental Causation," E.J. Lowe considers this way of responding to the exclusion problem.[134] (Where Lowe appeals to "non-coincidental causal overdetermination," I will instead take his account to be a rejection of overdetermination.

---

[132] In a footnote, Bennett notes that this claim is sometimes called the "exclusion principle" and that "not everyone emphasizes that aspect of the exclusion problem as much as they should." (493, fn.5)

[133] See, for example, Andrew Melnyk, *A Physicalist's Manifesto*. (Cambridge: Cambridge University Press, 2003). Alternatively, Stephen Yablo appeals to the relationship between *determinate* and *determinable*, in "Mental Causation" Philosophical Review 101, pp. 245–280. (1992)

[134] Lowe 2003, p.146-147

As I previously noted, the difference between positing nonredundant overdetermination and denying overdetermination is, for our purposes, insignificant.) He writes:

> This seems to me to be a perfectly fair objection on the part of the non-reductive physicalist…However, it may not be immediately apparent how an interactive dualist could hope to exploit the same sort of objection…How can the dualist maintain that systematic, non-coincidental causal overdetermination may be a widespread feature of situations involving mental causation, without conceding that mental events are ontologically dependent on physical events?

Because, as Lowe notes, the dualist denies that even a realization relation obtains between physical and mental events, she will need to find some other way of justifying the claim that mental causes, though coincident with physical ones, are nevertheless distinct and nonredundant. Lowe himself offers two very different ways of doing so; the first is a rejection of *exclusion*, the second is a bit more difficult to classify.

Before looking more closely at what Lowe has in mind, I think a brief digression is in order. After all, I have maintained that Lowe's first position is a rejection of *exclusion*, but Lowe himself does not make that claim. Instead, Lowe has presented his argument, alternatively, as a counterexample to a causal closure principle, and as the acceptance of systematic overdetermination. In "Causal Closure Principles and Emergentism," Lowe's proposal is clearly intended to show the inadequacy of a variety of causal closure principles. Indeed, Lowe makes no mention of *exclusion*, or of any other definition of (or sufficient condition for) overdetermination. There, when formulating the casual exclusion argument that he takes as his target, he writes that the argument has three premises:

> first, a physical causal closure principle; second, the claim—to which interactionist dualists are themselves committed—that at least some mental events are causes of physical events; and third, the claim that the physical effects of mental causes are not, in general, causally overdetermined.[135]

[135] Lowe 2000, p.572

The additional claim made by *exclusion*—that two simultaneous causes of a single event must be identical or overdetermining—Lowe does not make explicit.

In the later "Physical Causal Closure and the Invisibility of Mental Causation," Lowe amends the premises of the exclusion argument slightly.[136] There, he replaces the claim that "the physical effects of mental causes are not, in general, causally overdetermined" with the following "non-overdetermination principle:"

> (NOP) Most physical events *e* are such that, if *e* has a mental cause at time *t*, then *e* does not also have a wholly physical sufficient cause at *t* which is wholly distinct from that mental cause.

In other words, Lowe supplements the basic claim—there is no systematic overdetermination—with one that invokes a specific definition of overdetermination. This is significant, for it enables Lowe to reject *either* (a) the claim that overdetermination, as described, occurs, *or* (b) the claim that the scenario described, if it occurred, would in fact be overdetermination. That is, it is significant for precisely the same reason that *exclusion* is significant: it disambiguates two distinct responses to the causal exclusion argument.

Not surprisingly, then, Lowe adopts the second line of response. He rejects this new premise (NOP) by demonstrating a way in which the world could be such that most physical events *would* have both a mental and a physical sufficient cause, and yet neither cause would be redundant. For that reason, while Lowe's account does not explicitly take *exclusion* as its target, I nevertheless maintain that it is *exclusion* that is called into question by the scenario envisaged by Lowe. As we shall see, Lowe's argument essentially amounts to the claim that the physical world could be causally closed, and two

[136] Lowe 2003, p.137-154

simultaneous causes might nevertheless be distinct and nonredundant. In light of the distinction introduced by *exclusion*—one that, I suggest, is also at the heart of Lowe's rewritten premise—it seems the best way to understand Lowe's position is not as a criticism of a particular closure principle, nor as the acceptance of systematic overdetermination, but rather as a demonstration of the inadequacy of *exclusion*. If Lowe is correct, then *exclusion* fails to give a sufficient condition for overdetermination; two causes of a single effect might be distinct, simultaneous, and nevertheless nonredundant.

§3.2 E.J Lowe on Simultaneous Causation

Lowe's first attempt at defusing the causal exclusion argument begins with the question posed above. I include it here again, this time with Lowe's initial answer:[137]

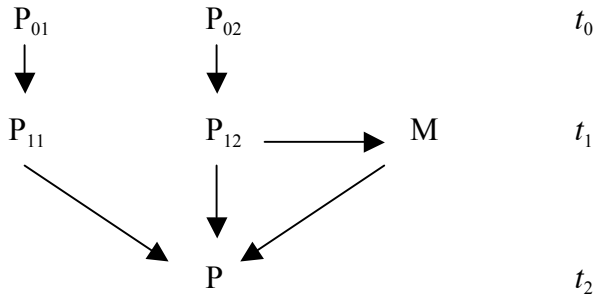
> How can the dualist maintain that systematic, non-coincidental causal overdetermination may be a widespread feature of situations involving mental causation, without conceding that mental events are ontologically dependent on physical events? The answer is remarkably simple: he may do so by maintaining that mental events, while not ontologically dependent on physical events, are *causally* dependent on them in certain ways. (147)

Suppose that some physical cause *p* suffices for a physical effect, *e*. In sufficing for *e*, *p* need not be the *immediate* cause of *e*, but might instead go through some intermediary which, itself, is the immediate cause of *e*. Suppose that this is the case, and that the causal intermediate between *p* and *e* is *m*—a mental event. If the causal relation between *p* and *m* is diachronic, then—unless there is some *additional* physical cause of *e* that occurs when *m* does—the scenario envisaged violates *closure*. If, on the other hand, we allow for the

[137] As noted, there are multiple formulations of this argument. In what follows, I refer to "Physical Causal Closure and the Invisibility of Mental Causation." (Lowe 2003)

causal relation between *p* and *m* to be a *synchronic one*—if, that is, we allow for simultaneous causation—then such a scenario could be compatible with the truth of *closure*.[138]

To illustrate what he has in mind, Lowe offers the following diagram (148). (The arrows represent the causal relation.)



On the proposed scenario, at time t_0, P_{01} and P_{02} jointly constitute a sufficient cause of P. Things are a bit more complicated at *t1*. There, according to Lowe, we find *two* sufficient causes of P. On the one hand, we have P_{11} and P_{12} which, together, suffice for P. On the other hand, we have M, which is itself a sufficient cause of P. P, then, has two distinct sufficient causes at a single time. If we grant *exclusion*, then P must be overdetermined.

Yet P is *not* overdetermined, according to Lowe, for P_{11} and P_{12} suffice for P *by going through M*, and successive causes in a single causal chain are not, ordinarily, taken to be in competition with one another. Nobody would suppose that the first and second domino in a chain are in danger of overdetermining the toppling of the third, for the first

[138] Lowe does not actually invoke *closure* in his statement of the exclusion argument. Instead, he appeals to the following causal closure principle: "(CCP) For any physical event *e*, if *e* has a cause at time *t*, then *e* has a wholly physical sufficient cause at *t*." (Lowe 2003 p. 141) Here, and in what follows, I will continue to appeal to *closure* rather than Lowe's CCP. I do this primarily for the sake of consistency. However, the stronger CCP entails *closure*. Clearly, if it is true that a physical event has a sufficient physical cause at every time at which it has *any* cause, then it has a sufficient physical cause at every time at which it has a *sufficient* cause. In light of this entailment, if something violates *closure*, it violates CCP as well. For this reason, I don't believe that I do Lowe any disservice by replacing his CCP with *closure.* True, consistency with *closure* does not entail consistency with CCP, but given the fact that (a) *closure* is the more widely held principle and (b) Lowe himself is committed to the falsity of CCP (see p.145), it will suffice for our purposes to show consistency with *closure*.

suffices for the toppling of the third *by causing* the second to fall. In the same way, suggests Lowe, if we allow for the simultaneous causation of a nonphysical cause, then an event can have multiple, distinct causes at a given time without thereby being overdetermined. For all we know, adds Lowe, this is precisely how mental causation occurs in the actual world.

The physicalist could, of course, simply deny the possibility of simultaneous causation. However, as Lowe notes, to do so would be to add an *additional* premise to the causal exclusion argument; neither *closure* nor *exclusion* precludes the possibility of simultaneous causation, nor does any other premise of the argument. It follows, then, that the present formulation of the causal exclusion argument fails. If there can be two distinct sufficient causes of a single event at a given time without that event being thereby overdetermined, then *exclusion* is false and the argument does not go through.

§3.3 Objections to Lowe's Simultaneity Account

Apart from simply denying the possibility of simultaneous causation, Lowe considers a series of potential objections to his proposal. In this section, I will examine what I take to be the most pressing of these objections. First, one might object to Lowe's account by posing the following dilemma:

> Either P_{11} and P_{12} need the help of M to bring about P, in which case they are not jointly sufficient for P, or else they do not need the help of M, in which case M is redundant. (149)

This objection has some intuitive weight. After all, at time t_1, M seems to be an essential component of P_{11} and P_{12}'s joint sufficiency for P. Lowe himself notes that "if M had not

occurred, then the conjunction of P_{11} and P_{12}, even if it had occurred, would not have sufficed to cause P." (149) It seems fairly straightforward, then, that $P_{11}$and P_{12} *alone* do not suffice for P

Fortunately for Lowe, there is an equally straightforward response to this objection. If one takes seriously the possibility of simultaneous causation, then the apparent tension between the joint sufficiency of P_{11} and P_{12} on the one hand, and the necessity of M on the other, dissolves. As Lowe writes, "This is just to say that P_{11} and P_{12} cannot bring about P 'immediately,' but only (in part) *via* an intermediate effect, M." (149) There is, however, no reason at all to suppose that the only genuinely sufficient causes are immediate ones. To return to the domino analogy, we typically find no tension between the sufficiency of the first domino for the toppling of the third, on the one hand, and the necessity of the second domino in this process on the other. Furthermore, Lowe writes:

> [I]f physical determinism is true, certain physical events in the early history of the universe were causally sufficient for various physical events occurring today, despite the fact that those early events were only able to bring about their present-day effects *via* very long chains of intermediate effects. (149)

Once we allow for the possibility of simultaneous causation, we allow for the possibility of there being a *causal* intermediary where there is no *temporal* intermediary. Again, one might simply reject the possibility of simultaneous causation, but to do so would require an additional argument. Absent some such argument, the fact that P_{12} and M are simultaneous does not suffice to show that they are in causal competition with one another; they might, instead, be distinct links on a causal chain.

Still, Lowe acknowledges that the scenario he has described has the following "interesting feature:"

> Any scientist who was to examine that situation by empirical means, but who was restricted by his means of investigation to observing only purely physical events and causal relationships, would quite naturally come to the conclusion that the physical event P had a complete and wholly physical causal explanation…Such an investigator would notice no 'gaps' in the physical causation of P..." (150)

If mental causation occurs in the way proposed by Lowe, then it is invisible. We should not expect to find—indeed, should expect *not* to find—evidence of it in the physical sciences. In light of these considerations, Lowe asks, might not the scientist be entitled to reject either the causal efficacy or the irreducibility of M? If there is a chain of physical causes that is explanatorily sufficient for P, then shouldn't we conclude that this chain is, itself, *causally* sufficient for P as well?[139] If so, then doesn't it follow that M is either epiphenomenal or, in one way or another, reducible to some physical event?

"The answer," writes Lowe, "is clearly 'No,' because the situation depicted in the diagram rules out all of these options and yet is metaphysically perfectly possible." (151) Because there is a possible scenario on which mental causation is both *real* and *invisible*, the mere invisibility of a proposed mental cause cannot suffice to undermine one's justification in affirming its efficacy. Indeed, what Lowe's diagram shows is that the reality of mental causation is perfectly compatible with the *appearance* of an explanatorily closed physical system. That is to say, the fact that we can construct a gapless explanatory history for some physical event does not suffice to show that our construction is in fact a *complete* explanatory history. To return to the fictitious investigations of Lowe's scientist, Lowe notes that "The causal explanation of P in wholly physical terms *would* in fact be incomplete, of course, but it would not *appear* to

[139] By claiming that there is a chain of physical causes that is explanatorily sufficient for *p*, I do *not* mean to imply that there is a purely physical *complete* explanation of *p*. I mean only that the explanation would be gapless, and thus would appear to be complete.

be incomplete…" (151) It follows, then, that the mere invisibility of a mental cause does not suffice to ground either its elimination or reduction.

Might there be other reasons in favor of eliminating (or reducing) an invisible cause? On this point, Lowe briefly notes three "stock objections likely to be raised by physicalists." (151) First, the physicalist might reject Lowe's account for *economical* reasons. After all, if we can explain any given physical event without appealing to irreducibly mental ones, then doesn't Ockham's Razor compel us to do so? The answer, according to Lowe, is that "we just have no right to suppose that reality operates along the most 'economical' lines—that every effect is always brought about in the simplest possible way." (152, fn 10) This is particularly true in light of the fact that the dualist has independent reasons for affirming the existence of irreducible mental states.

This consideration takes us to the second objection: isn't it *ad hoc* to postulate irreducible mental states for which we have no physical evidence? Lowe responds:

> This may be a fair objection to raise against some forms of 'panpsychism.' But given that we do, where human brains are concerned, have reliable testimony confirming the occurrence of mental events, which at least *seem* to be neither identical with nor 'realized' by brain events, there need be nothing *ad hoc* and unprincipled about postulating that these events are precisely what they seem to be, namely, ontologically 'additional' non-physical events. (152)

Much has been written about the apparent irreducibility of qualitative mental states.[140] Without getting into the details here, this much is true: questions of whose account is *ad hoc*, like questions of ontological economy, depend largely on what body of evidence one chooses to privilege. If, to borrow a phrase from Chalmers, we "take consciousness seriously," then there is nothing either ontologically gratuitous or *ad hoc* about accepting

[140] Here I refer the reader to: Chalmers 1996, Kim 2003, Kripke 1980 and to Chapter One of this dissertation.

at face value apparently irreducible mental states. The abundance of first-person testimony regarding these states can suffice as a plausible defense of, at the very least, the *possibility* of their existence.

Still, the physicalist might say, these mental states are certainly *mysterious*, and "inexplicably at odds with what we have discovered about biological evolution." (153) Evolution is, from what we know, a gradual process. Nonphysical states could not *gradually* develop out of physical ones, for they are neither wholly nor partly composed of physical states; no minor change to a physical state could render it nonphysical. If, at some point in our evolutionary history, mental events came to be, that must mean that "a wholly new kind of event suddenly sprang into existence." (153)

Lowe offers two responses to this third objection. First, we don't actually *know* that evolution has always been gradual. He writes, "it is widely disputed whether all biological evolution is in fact gradual in character." (153) Second, (and, it seems to me, more to the point) we are considering the possibility of *nonphysical* states. As such, "there is no reason to expect their historical provenance to be governed by principles of evolutionary biology." (153) Biological evolutionary history may be able to tell us how *biological* entities have developed. Absent the claim that *all* entities are biological entities, we should not expect it to be able to tell us how *all* entities have developed. The committed dualist, then, should see no tension between the nonphysical nature of irreducible mental states and the gradual nature of biological evolution; the latter need have no bearing on the development of the former.

§3.4 Closure: Causal vs. Explanatory

If Lowe is correct about the possibility of simultaneous physical-to-mental causation, then he seems genuinely to have provided a counterexample to *exclusion*. Successive events in a causal chain are not generally taken to be in competition with one another, and it's hard to see why the mere simultaneity of the events should change that fact. (Unless, that is, the very possibility of simultaneous causation is disputed but, again, that would require an additional argument. The question here is whether or not simultaneous causal successors *if possible* would be overdetermined.) If Lowe's diagram illustrates a metaphysically possible scenario, then it demonstrates the possibility of an event's having two distinct sufficient causes at a single time without thereby being overdetermined. It thus demonstrates the failure of *exclusion* to serve as a sufficient condition of overdetermination.

At the same time, there is something a bit fishy about Lowe's response to the causal exclusion argument. After all, this response is supposed to be one that is compatible with the truth of *closure*.[141] Yet, if the physical domain really is causally closed, then it seems that a scenario of the sort described *shouldn't* be possible. To see why, note that Lowe's mental cause is supposed to be ontologically distinct from any physical cause *and* to make a genuine *difference* in the physical world. Recall that, according to Lowe, "if *m* had not occurred, then the conjunction of P_{11} and P_{12}, even if it had occurred, would not have sufficed to cause *p*." (149)[142] For that reason, all

[141] Again, it is explicitly intended to be compatible with the stronger CCP, but as CCP entails *closure* it should, of course, also be compatible with *closure*.

[142] Of course, if P_{12} suffices for *m*, then P_{11} and *p*12 could not occur absent *M*. It remains the case, however, that *M* is *necessary for P*, and that is the relevant point in this context.

appearances of completeness aside, "the causal explanation of *p* in wholly physical terms *would in fact be incomplete*." (151) (emphasis added) On Lowe's diagram, then, P is a physical event for which there is no complete, wholly physical explanation. In order to explain P—and in order to explain *any* mentally caused physical event—we need to go *outside* of the physical domain and appeal to a nonphysical, mental event. This is, of course, what the interactive-dualist believes to be true anyway, and is not itself an objection to Lowe's account. It should, however, give us pause as to the question of whether or not the scenario envisaged by Lowe is compatible with the causal closure of the physical domain.

Perhaps, then, the physicalist ought to reject *closure* in favor of some other, better causal closure principle. After all, aforementioned worries aside, Lowe's diagram *has* been shown to be compatible with *closure*. If it is not compatible with the physical domain's being causally closed, then clearly *closure* does not do the job for which it is intended. Alternatively, because *closure* is impugned by Lowe's argument only if the possibility of simultaneous causation is granted, perhaps the physicalist ought simply to reject the possibility of simultaneous causation. In either case, the causal exclusion argument will need to be reformulated if it is to preclude the possibility of a scenario such as the one envisaged by Lowe.

In section 5, we will consider a response to the causal exclusion argument that, according to Lowe, defeats even those formulations that *both* prohibit simultaneous event causation and appeal to a stronger closure principle. Lowe's second argument is quite different from the first, but the two are alike in one way: both have the result that some physical event is dependent for its occurrence on a nonphysical mental cause. On both of

Lowe's accounts, then, causal closure and what I have called "explanatory closure" come apart; the former holds for the physical world, but the latter does not.

§4.1 Some Thoughts on Freedom

Before considering Lowe's second proposal, it is worth noting one feature that his first account has in common with that of Mills: neither is compatible with the libertarian conception of free agency. The scenario envisioned by Lowe, like the scenario envisioned by Mills, might leave room for mental acts, but it certainly does *not* leave room for *free* mental acts—not, that is, if by "freedom" we mean "libertarian freedom."

In *An Essay on Free Will*, Peter Van Inwagen offers the following definition of (libertarian) free will:

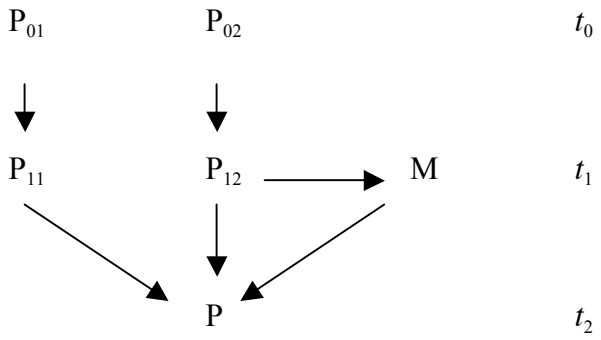
> When I say of a man that he "has free will" I mean that very often, if not always, when he has to choose between two or more mutually incompatible courses of action—that is, courses of action that it is impossible for him to carry out more than one of—each of these courses of action is such that he can…carry it out.[143]

To say that a person is free in the libertarian sense, then, is to say that she is free to choose between more than one course of action. Of course, as a general rule, a person is not free to do the impossible. I do not ever get to choose between flying down the street and walking down the street, nor am I free to breathe, unassisted, underwater. (There might be *some* circumstances in which we would want to say that a person freely chose what was, in actuality, an unavoidable course of action, but if so, then surely these

[143] Van Inwagen1983, p.8

situations are the exception.)[144] Given that freedom is constrained by possibility in this way, if an event is the causal result of a free choice, then it will generally be the case that it was possible for that event *not* to have occurred when it did. An event that is freely caused in the libertarian sense is an event that *might not have occurred*.

In light of this fact, consider, once again, the scenario suggested by Lowe: At some time (t_1), a physical event (P_{12}) causes a mental event (M)—bringing about M's *simultaneous* occurrence—and M subsequently causes the physical event P to occur at t_2.



If *this* is how mental causation works, then there are no freely chosen acts; the situation depicted above is incompatible with libertarian agency.

To see why, note that the mental cause, M, is the result of a sufficient cause. If P_{12} suffices for M, then M cannot have failed to occur when it did. Furthermore, because P_{11} and P_{12} (by going through M) jointly suffice for P, P likewise cannot have failed to occur when it did. But if this is true for all mentally caused effects, then no mental cause ever brings about anything other than that which was determined by antecedent circumstances. According to Lowe's proposed scenario, a mental event can play a causal role in the physical world *only* insofar as that event is a necessary causal link in a mostly-physical

[144] Suppose, for example, that I must choose between looking to my right or looking to my left. It might turn out to be the case that, were I to choose to look to the left, my neck muscles would spasm in such a way as to prevent such a movement from occurring. Still, should I choose to look to the right, we might nevertheless maintain that I was free in doing so.

causal chain—its own occurrence, and its own effect, wholly determined by prior events. A mental event of *that* sort cannot suffice for libertarian freedom.[145]

Mills' account is likewise incompatible with libertarian agency. If every mentally caused event is *also* the result of an unbroken chain of physical sufficient causes, then there is but one event that could have occurred at that time. There is a sense in which this objection is just a restatement of the one raised against Mills in section 2.2. If Mills is correct about the way that mental causation works in the actual world, then no mental cause ever results in a deviation from the causal path determined by physical antecedent circumstances. The mental cannot make a difference in the world, even if we wish nevertheless to say that the mental is—in some sense—causally efficacious in the physical world.

If a substance dualist chooses to respond to the causal exclusion argument along either of the lines so far considered, then—in doing so—she closes the door to libertarian agency. This is a startling conclusion. After all, while a dualist need not *affirm* libertarian freedom, it would be quite a disadvantage to learn that she must *deny* it. But if she chooses to respond to the exclusion problem in one of the ways we have been discussing, then it looks like she must do just that. For if mental causation is merely overdetermining causation, or if mental causes are links in an unbroken chain of sufficient causes, then there is simply no room for freedom.

[145] This would come as no surprise to Lowe, who himself notes this feature of his first account. It is, nevertheless, worth making explicit here. (Lowe 2003) p.148

§4.2 Freedom, *Closure*, and Completeness

In the foregoing discussion, I have taken Jaegwon Kim's *closure* to be the definitive statement of the causal closure of the physical world. It is not, however, the *only* statement of causal closure, and it is worth asking whether the alleged incompatibility under discussion arises from *closure* itself, or from causal closure broadly understood. In what follows, I hope to show that *closure* is not the problem.

In doing so, I will focus not on the differences among competing statements of causal closure, but rather on what they have in common—the *spirit*, as it were, of causal closure.

The causal closure of the physical domain is sometimes taken to be synonymous with a similar thesis, the completeness of physics, and indeed the two go hand-in-hand.[146] Like causal closure, there are a variety of ways of stating the completeness of physics, but the general idea is straightforward: physics is *complete* in that it is causally, and explanatorily, self-sufficient. There are no physical events for which it is necessary to appeal to something *outside* of the domain of physics in order to account for its occurrence.[147] It is thus easy to see how closure and completeness are related, for if the physical world is causally closed, then physics—the study of the physical world—will be a complete (or completable) science. If, on the other hand, the physical domain is *not* causally closed, if it is instead susceptible to outside interference, then—for at least some physical events—physics alone will never be able to tell us the whole story. For this

[146] (Strictly speaking, it seems the two cannot actually be synonymous; the former refers to a domain of physical objects, the latter to a branch of science.)

[147] Without getting into a detailed discussion on the metaphysics of explanation, it should be noted that not all advocates of causal closure would agree that the physical world is *explanatorily* closed. Some nonreductive physicalists, for example, would claim that certain *features* of a physical event cannot be explained without reference to irreducibly mental properties. Still, the *occurrence* of the event itself, at least under some physical description, must be explainable in physical terms if the physical is causally closed—by reference to the event's physical causal history—and that is all that I mean to say.

reason, the causal closure of the physical world, properly understood, ought to entail the completeness of physics.[148]

Similarly, the completeness of physics—if demonstrable—would give us good reason to believe in the closure of the physical domain.[149] In "Some Evidence For Physicalism," Andrew Melnyk offers the explanatory success of physics as evidence of the truth of causal closure. According to Melnyk, the best evidence for the truth of causal closure can be found in, or at least gleaned from, physics textbooks:

> Although the claim that the physical is causally closed is not explicitly stated in physics textbooks, it may nonetheless be inferred from claims that *are* explicitly stated in physics textbooks. According to the textbooks, then, contemporary physics has succeeded in finding sufficient physical causes for physical effects of very many kinds; and it has found no physical effects at all for which it is necessary…to invoke non-physical causes. But current physics' success to date in finding that *many* physical events have sufficient physical causes provides inductive evidence that *all* physical events, including *both* unexamined physical events *and* examined-but-as-of-yet-unexplained physical events, have sufficient physical causes. (160-161)

We will evaluate the strength of Melnyk's inductive argument for causal closure in the next chapter. What is of central importance here is that we note the close connection between the *completeness* of physics, on the one hand, and the *closure* of the physical domain on the other. As Melnyk notes, the fact, (if, indeed, it is fact) that physics can provide us with a complete causal story about the world *without ever having to appeal to the nonphysical* serves as evidence in favor of the causal closure of the physical domain.

In the same way, were it to be the case that physics could *not* provide such a story, but instead had to occasionally appeal to something *outside* of the physical domain for a

---

[148] Again, "completable" is perhaps the better term here. Surely current physics is not *yet* complete, if by complete we mean wholly explanatorily sufficient. Still, if closure is true, then physics ought to be *in principle* complete, if not in practice.

[149] As we shall see in Chapter Four, whether or not completeness actually *entails* causal closure is less clear.

complete causal history of some physical event, this would in turn serve as evidence *against* causal closure. Suppose, for some physical event *p*, that: (a) *p* has an explanatorily sufficient, gapless causal history, but (b) *p* lacks an explanatorily sufficient, gapless *purely physical* causal history. Then, assuming that physics cites only physical causes in its explanations, physics could not give a complete causal explanation of *p*—*despite* the fact that *p* is wholly explicable. Physics would thus be incomplete. It is, it seems to me, but a short step from this incompleteness to the conclusion that the physical world is not causally closed. After all, regardless of the *details* of one's definition of causal closure, this much seems true: if the nonphysical can make a *causal difference* in the physical world, and can serve to explain the occurrence of a physical event that the physical causal history, alone, cannot, then surely the physical domain is not causally closed. A causally closed system cannot be susceptible to this kind of causal *interference*—or what could it mean to say that it was causally *closed*?

We are now in a position to see why it is causal closure generally understood, and not *closure* in particular, that poses a problem for the libertarian interactionist. The scenario just proposed—the existence of a physical event the occurrence of which cannot be explained without reference to a nonphysical cause—is precisely the sort of scenario to which the libertarian dualist must be committed. Simply stated, for there to be *free*, *nonphysical*, *causal agents*, the causal effects of which can be found in the physical world, these nonphysical agents must be able to *make a difference* in the physical world. They must be able to bring about events that might not have been brought about, to cause things that might not have been caused. They must, that is, *interfere* in the physical domain. It is difficult to conceive of an account of causal closure that would allow for such interference, for it is difficult to see how an interfered-with causal order could be a

closed one. The claim that there *are* such agents, therefore, essentially amounts to the claim that physics is not complete, and that the physical domain is, likewise, not causally closed.[150]

Perhaps there is a way of accommodating substance dualism, libertarian freedom, and the causal closure of the physical world, but if so I can't see it. E.J. Lowe and Eugene Mills both succeed in demonstrating the compatibility of *closure* and interactionist dualism, but they do so at great cost. A commitment to interactionist dualism ought not to commit the dualist to the denial of libertarian freedom. If this unfortunate commitment can be avoided, then it ought to be avoided. In the next chapter, I will argue that an interactionist dualist need not accept the causal closure of the physical world, and that—in light of these considerations—a rejection of *closure* is the simplest and most advantageous response that a dualist can make to the exclusion argument. Before doing so, however, I will consider one more attempt at reconciling *closure* and interactionist dualism. This final account, unlike the first two, purports *also* to be compatible with libertarian freedom.

§5. A Third Alternative: E.J. Lowe on Fact Causation

In "Non-Cartesian Substance Dualism and the Problem of Mental Causation," Lowe offers an alternative response to the causal exclusion argument—one that,

[150] Really, the trouble is not that physics is *presently* incomplete, but that it is not *completeable*. Assuming that there remains much that is as of yet unbeknownst to physicists, physics is of course incomplete. The problem with free, nonphysical agents is that their existence ensures that physics will remain incomplete, no matter the progress of the physicist.

according to Lowe, is compatible both with *closure* and with libertarian agency.[151] Once again, Lowe accepts the statement of causal closure that he takes the physicalist to be working with, and he accepts the premise that mentally caused events are not systematically overdetermined. (He would claim only that they are not *redundantly* overdetermined.) As such, it might seem that this second response, like the first, is a rejection of *exclusion*. This, however, is not quite right. Instead, Lowe rejects the very first premise of the causal exclusion argument: despite the reality of mental causation, Lowe maintains that no nonphysical mental event is ever the cause of a physical event. As such, there is no causal competition between physical events and nonphysical ones.

Lowe's second response to the exclusion argument goes roughly as follows: In order to fully understand any instance of human action, we need to appeal to two distinct kinds of causation. On the one hand, there is purely physical event causation; on the other, there is mental causation. Mental causation, argues Lowe, is not a subspecies of event causation, for it is *intentional* where event causation is "blind." Instead, mental causation is best understood as *fact* causation. Furthermore, mental *fact*-causes and physical *event*-causes do not compete with, but rather compliment one another. Where an intentional mental cause determines what *kind* of event occurs in an instance of mental causation, there will always be a physical sufficient cause to determine the *particular* features that comprise the token event that in fact occurs.

Consider the following example: Suppose that I am thinking about raising my hand in class. I deliberate, and at some point in time (*t*) I make the *choice* (C) to raise my hand. As a result, my arm rises shortly thereafter (AR). Now, if *closure* is true, then at *t* there is a sufficient physical cause of my arm's rising. Call that cause P. How is it that my

[151] *Erkenntnis* (2006) 65:5–23

(nonphysical) act of choice can suffice for an event that has a sufficient, purely physical cause, without thereby overdetermining that event? The answer, according to Lowe, is that each is causally responsible—causally *sufficient*—in its own way. Where the mental cause determines what *kind* of event occurs—for example, an arm-rising rather than an arm-resting—it is impartial with respect to which of the many possible instances of that kind actually occurs. The smoothness and rapidity of the arm-rising, for example, or the precise time at which it commences, are features for which the fact-cause is not responsible. I do not choose to raise my arm at any particular angle, velocity, or even at any precise point in time; I simply choose to raise my arm. These *particular* features—and the particular event that exemplifies them—are determined by the purely physical sufficient cause.

If this account is correct, then for any mentally-caused event, there will be aspects of the event that neither the mental nor the physical causal history could itself suffice to explain. It is important to note the symmetry with respect to explanatory failure that Lowe posits. This is not an account on which the mental cause simply allows for a *redescription* of an event, the occurrence of which is wholly explicable in physical terms. On the contrary, Lowe maintains that an act of choice is, in general, *free* in the libertarian sense; it does not itself have a sufficient cause. In the example above, when I chose to raise my arm at *t*, I might instead have chosen *not* to raise my arm. Had I made that choice, according to Lowe, *my arm would not have risen*. More importantly, my choosing *not* to raise my arm really could have happened; nothing in the physical antecedent circumstances determined, or prevented, my making either choice. On this proposal, then, the physical cause cannot fully explain the arm-rising, for it cannot explain why it was an arm-rising, and not an arm-resting, that occurred. Likewise, my choice to raise my arm

cannot explain the smoothness with which the arm rose, or the precise moment at which the arm-rising commenced. For a complete explanation of AR, we need *both* the physical cause (P) and the nonphysical choice (C).

There are, of course, a number of questions that must be addressed in order to evaluate this account. Before considering these questions, however, we ought first to get clear on the following: how, exactly, is this a response to the causal exclusion argument? Consider again the following formulation of the argument:

(1) Suppose mental event M causes physical event P at *t*. (*for reductio)*
(2) P has a sufficient physical cause at *t* as well, call it P*. (*closure*)
(3) M is not identical with P*. (substance dualism)
(4) P cannot have more than one complete cause at *t*—unless this is a case of genuine overdetermination. (*exclusion*)
(5) This is not a case of genuine overdetermination.
(6) Then either P* or M is the cause of P, but not both. ((1)-(5))
(7) P*, not M, is the cause of P.
(8) If M causes P at *t*, then M does not cause P at *t*. ⊗

If Lowe is correct, then the dualist can affirm the reality of mental causation while denying the truth of premise (1). In its place, the dualist might offer the following premise:

(1*) Suppose mental fact M causes physical fact P at *t*.

Unless the physicalist is prepared to endorse the claim that the domain of physical *facts* is also causally closed, the argument cannot proceed from (1*).

Might the physicalist simply assert the causal closure of the domain of physical facts? Lowe considers this, but notes that, in doing so, the physicalist would abandon all pretense of having an *argument* against the possibility of interactive substance dualism. Instead, this would amount to a dismissal of the possibility of nonphysical causation at the outset. Lowe writes,

> To assert that any cause of anything physical must itself be physical is equivalent to asserting that no cause of anything physical can be nonphysical, which directly contradicts the interactive dualist's claim that something physical may have a non-physical cause. A 'causal closure argument' that appeals to a principle of causal closure which is itself inconsistent with interactive dualism amounts, in effect, to nothing more than this: P, therefore not not-P. Hence, it is in the physicalist's own interest not to appeal to a causal closure principle that is so overridingly strong as this. (21, fn28)

Unlike *closure*, which allows at least for overdetermining or simultaneous mental causes, a closure principle strong enough to rule out mental fact causation would be too strong to do the work for which it is intended. It would not, therefore, serve to bolster the causal exclusion argument. On the contrary, it would render the argument unnecessary.

§5.2 Objections to Lowe's Second Approach

Suppose that Lowe's proposed scenario is metaphysically possible. One might nevertheless ask, as we did with Mills's account, whether or not it is *plausible*. More specifically, the physicalist is likely to raise many of the same "stock objections" to this account as she did to the first: isn't it, after all, *ontologically extravagant, mysterious, and ad hoc* to posit nonphysical fact causes in addition to the physical event causation to which we are already committed? It is, however, none of the three if we have antecedent reasons to believe both in the *causal efficacy* of our choices and in the *irreducibility* of those choices to physical events. As for the first, Lowe would be well-served by an appeal to the oft-cited words of Fodor:

> If it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying...if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world. [152]

[152] Jerry Fodor, *A Theory of Content and Other Essays*. (Cambridge, Mass: Bradford Book/MIT Press, 1990) p. 156

At the very least, the dualist has *prima facie* reasons to affirm the causal efficacy of choice.

As for the second, Lowe himself argues for the apparent irreducibility of the act of *choosing*, or the choice that results from this act, to any single or composite physical event. If contemporary neuroscience is correct, he notes, then physical causation in the brain is, in general, the result of a great confluence of causal chains, many of which are not obviously related to one another. Mental causation, and in particular the act of *choice*, is rather different. He writes,

> What seems plausible is that if we were to trace the purely bodily causes of any peripheral bodily event, such as the movement of my arm on a given occasion, backwards in time indefinitely far, we would find that those causes ramify, like the branches of a tree, into a complex maze of antecedent events in my nervous system and brain – these neural events being widely distributed across large areas of those parts of my body and having no single focus anywhere, the causal chains to which they belong possessing, moreover, no distinct beginnings…And yet, my mental act of decision or choice to move my arm seems, from an introspective point of view, to be a singular and unitary occurrence which somehow initiated my action of raising my arm. (11-12)

Neural causes and mental causes seem quite different, and it thus "seems impossible" to identify a choice with any individual or composite neural event. (12)[153] The dualist, then, need not worry about charges of extravagance, or of holding an *ad hoc* position. She can, instead, appeal to independent evidence in favor of the existence of ontologically distinct, causally efficacious mental states.

What, though, of the relation of these mental states to physical ones? Can we reconcile the efficacy of a nonphysical cause with what we know—or at least believe—

[153] See (Lowe 2006) pp11-15 for a more robust defense of this claim. Lowe gives a much more extensive defense of the irreducibility of choice to neural states than the one I have here provided. In particular, he notes the difference in counterfactuals sustained by *p* and C as further evidence of their distinctness..

about the physical world? One objection along these lines that has likely already struck the reader is the following: Lowe's proposal seems not to be compatible with the claim that physical event-causation is deterministic. If *p* is a sufficient cause of AR, as Lowe maintains, but *p* could have failed to cause AR—if, for example, I chose not to raise my hand at *t*—then *p* suffices for, without determining, AR.[154] Lowe anticipates this charge, and responds as follows:

> Maybe so. But in view of the developments in quantum physics during the 20th century, we now know that physical causation is not in fact deterministic, so the objection is an idle one and can safely be ignored. (19)

Because physics tells us that there is genuine indeterminacy at the quantum level, Lowe is untroubled by the indeterministic picture of physical causation to which his account seems committed.

Much more can be said about the deterministic, or indeterministic, nature of physical causation. In particular, Lowe's account would be well-served by a defense of the claim that *macro*-indeterminism is a real feature of this world, in addition to the micro-level indeterminism to which quantum physics commits us. However, the deterministic nature of physical causation is not a premise of the causal exclusion argument. For that reason, we can bracket these concerns for the time being.

The causal closure of the physical world, on the other hand, *is* a premise of the exclusion argument. Furthermore, Lowe maintains that his account is consistent not only with *closure*, but with the following, stronger statement of causal closure:

---

[154] Might we not just say that, had I made a different choice at *t* (C*), *p* would not have obtained? If so, then we need not say that *p* failed to cause AR, only that some other physical cause, *p**, obtained instead of *p*. I don't think this will avoid the problem, however. For *p*, like AR, has some immediately prior sufficient cause—call it *p**. If, at *t*, *p* fails to obtain, then *p** must have failed to cause *p*. The problem cannot be avoided by taking a step back, so to speak. If both C and C* were metaphysically possible at *t*, and if AR has a sufficient physical cause, then a sufficient physical cause must not need to be a deterministic cause. (At least, that's how it seems to me. Lowe does not state the case as strongly as I have.)

> *Closure**: the domain of physical events is causally closed, in the sense that no chain of causation can lead backwards from a purely physical effect to antecedent causes some of which are nonphysical in character. (11) ("Closure*" is my own terminology, not Lowe's.)

*Closure** is stronger than *closure* for the following reason. *Closure* claims that no physical event can have a sufficient nonphysical cause at a time, *t*, unless there is *also* a physical sufficient cause at that time. It thus allows for nonphysical links in an otherwise physical causal chain, provided that those links coincide with physical ones. *Closure**, in contrast, disallows even this. Instead, it maintains that there can be *no* nonphysical links on a physical causal chain; even a redundant nonphysical event-cause is disallowed by *closure**.[155]

If there are physical events, like AR, that have genuinely efficacious, nonphysical causes, then doesn't that mean that both *closure* and *closure** are false? According to Lowe, it does not. First, recall that, on Lowe's scenario, AR has a sufficient physical cause that occurs precisely when C does. *Closure*, then, is not violated. But neither is the stronger principle, *closure**, for any physical chain of causes leading back from AR—which is, after all, a physical event—will lead only to additional *physical* events. An event causal chain is comprised only of physical events; the mental fact cause, the choice, will not be a part of that causal chain. As such, it will be *true* that "the domain of physical events is causally closed, in the sense that no chain of causation can lead backwards from a purely physical effect to antecedent causes some of which are nonphysical in character." (11)

[155] One might, at this point, wonder why *closure** is not too strong for the purposes of the causal exclusion argument. After all, it rules-out the possibility of mental-to-physical event causation. It does not, however, rule-out the possibility of mental causation *tout court*, for it allows for the possibility of mental fact-causation. For Lowe, this is enough. A dualist who is committed to mental *event* causation, however, might not be so amenable to the physicalist's appeal to *closure**.

§5.3 Reflections on Lowe's Second Alternative: Explanatory and Causal Closure Revisited

In addition to the objections that Lowe himself considers, questions remain about the tenability of Lowe's second response to the exclusion argument. One striking feature of this account is that no mental event or fact, no *choice*, is ever the cause of a physical event. We have, of course, noted this, but it is a bit difficult to see how to reconcile this fact with the claim that choices are efficacious in the physical world. If my choice to raise my arm never causes an arm-rising, then *how* does it cause it to be true that an arm-rising occurs? How does a choice cause it to be true that an event of a certain type occurs, without ever causing an event of that type?

If I understand Lowe correctly, then he would respond along the following lines. For any physical event that is the result of a choice, there is a sufficient, yet indeterministic, physical cause. The mental cause, the *choice*, determines which of the possible effects actually follows from the physical cause—perhaps by fixing the chances of the effect. In this way, though the choice does not cause the effect itself, it causes it to be the case that this effect, and not some *other* possible effect, actually follows from the physical cause.

Perhaps an account of this sort can be worked-out. There is, however, the following worry: there still seems to be a physical event that has as its immediate cause a choice. If P is the physical cause of AR, but P only causes AR because my choice *determined* that P would cause an arm-rising, then it seems there is a physical event (E) that consists of P's causal powers being determined. We might not wish to appeal to

causal powers, we might choose instead to speak of "fixing chances," but the point remains: if the choice is responsible for the fact that *p* causes AR, then it seems as if there is a physical event lurking somewhere just prior to (or perhaps simultaneous with) P's causing AR. If there is, then what might the cause of *that event* be? Not another physical event, for *given the physical history alone*, *p* might just as well have caused an arm-resting as an arm-rising. No, it seems that the choice—a nonphysical, mental cause—is the obvious candidate. If so, then Lowe's second proposal is not consistent with *closure**, or with *closure*, after all.

In light of the preceding discussion of libertarian freedom and causal closure, this ought not to be surprising. Either the nonphysical can make a difference in the physical world—can make a physical event occur that *would not otherwise* have occurred—or it cannot. If it can, then physics is not complete and the physical world cannot plausibly be deemed "causally closed."[156] If it cannot, then libertarian freedom is not a part of our world.

Concluding Thoughts

The causal exclusion argument is intended to show the necessity of reduction for mental causation. In this respect, the argument fails. As we have seen, a dualist who wishes to affirm mental causation might simply follow Mills in accepting the systematic

[156] Note that I say "plausibly," and not "possibly." If Lowe is correct, then the physical world is not *explanatorily* closed, but it is *causally* closed. As I have said, I don't see how the details of this account can be worked out. I don't see, that is, how a nonphysical choice can bring-about the kind of event that occurs without, at any point, *causing* a physical event to occur. If this account can be worked out, however, then the incompleteness of physics and the lack of explanatory closure would not entail the causal closure of the physical world.

overdetermination of mental causes. Alternatively, she might sit the existence of "invisible" mental causes that are causally antecedent to, but simultaneous with, some physical cause, as Lowe does. Neither option is without challenges, but both constitute viable responses to current formulations of the causal exclusion argument. Finally, she might instead reject the first premise of the causal exclusion argument and maintain, again with Lowe, that mental causation is not event causation at all, but is another form of causation entirely. Whether she rejects the claim that there is no systematic overdetermination, or rejects the definition of overdetermination put forth in *exclusion*, or rejects the event-causal nature of mental causation, the dualist has options.

The question, of course, is whether or not any of these options are particularly *good* ones. While it has been shown that the causal exclusion argument does not succeed in demonstrating the *necessity* of reduction for mental causation, it would be nice if the dualist had a response that was *palatable*. Both Mills's account and Lowe's first account require the dualist to abandon the possibility of libertarian agency. If I am correct in my assessment, Lowe's second account does the same. All *three* accounts involve adopting a fairly unconventional position with respect to some philosophical position, whether it be simultaneous event causation, systematic overdetermination, or the distinction between *fact* causes and *event* causes. The question, then, is whether or not it's worth all that just to avoid rejecting the causal closure of the physical domain. I don't think it is. Instead, in Chapter Four, I will argue that the dualist ought rather to stop trying to render interaction compatible with causal closure.

CHAPTER FOUR:

CHOOSING NOT TO WORRY ABOUT CLOSURE

In Chapter Three, I argued that the causal exclusion argument fails. In light of the possibility of systematic overdetermination, as shown by Eugene Mills, and the possibility of simultaneous mental causation, as shown by E.J. Lowe, it is clear that the argument fails to demonstrate the necessity of reduction for mental causation. At the same time, I suggested that neither Mills nor Lowe has provided the substance dualist with a palatable theory, for neither theory is compatible with the possibility of libertarian agency. In what follows, I hope to show that the lengths to which Mills and Lowe have gone in order to preserve the truth of *closure* are unnecessary, and that the dualist who wishes to respond to the exclusion argument ought rather to focus her attention on *closure* than on *exclusion*.

In §1, I examine the implications of a rejection of *closure* on the causal exclusion argument. There I will show that, without *closure*, the argument simply goes away. Nevertheless, the rejection of *closure* is an unpopular position (to say the least). For this reason, I devote §2 to the question of evidence *in favor* of causal closure. In §3, I will

raise what I take to be the most significant challenge to any statement of the causal closure of the physical: Hempel's Dilemma. Accordingly, §4 and §5 will treat physicalist responses to the dilemma, the first and second horn of the dilemma respectively. Finally, in §6, I will argue that there is *no* causal closure statement that can adequately ground a causal exclusion argument against interactionist dualism. For that reason, the dualist ought *not to worry* about closure. Depending upon how one defines "physical," the resulting causal closure principle will either be (a) false, (b) compatible with interaction, or (c) obviously question-begging.

§1. *Closure* & The Causal Exclusion Argument

Suppose, as I have argued, that interactionist substance dualism cannot plausibly accommodate both libertarian freedom and *closure*, and that the dualist ought rather to affirm the former than the latter.[157] What follows if the dualist chooses to reject *closure*? Well, for one thing, the causal exclusion argument loses its argumentative force. Consider, once again, the argument:

(1) Suppose mental event M causes physical event P at *t*. (*for reductio*)
(2) P has a sufficient physical cause at *t* as well, call it P*. (*closure*)
(3) M is not identical with P*. (substance dualism)
(4) P cannot have more than one sufficient cause at *t*—unless this is a case of genuine overdetermination. (*exclusion*)
(5) This is not a case of genuine overdetermination.
(6) Then either P* or M is the cause of P, but not both. ((1)-(5))
(7) P*, not M, is the cause of P.

[157] I do not mean to say that all interactionist dualists must be committed to a libertarian conception of freedom, or to the claim that such freedom obtains in the actual world. However, given the choice between (a) affirming *closure* and denying the *possibility* of libertarian agency, and (b) denying *closure* and allowing for the possibility of libertarian freedom, I think the latter is the clear choice. I will, of course, defend this claim in the course of this chapter.

(8) If M causes P at *t*, then M does not cause P at *t*. ⊗

According to the proponent of the causal exclusion argument, nonphysical mental causation must either be (a) systematically overdetermined, or (b) not, in fact, nonphysical after all. Absent *closure*, however, the conclusion simply doesn't follow.

By rejecting *closure*, the dualist is free to reject premise (2) of the argument. Without the assumption of a rival physical cause of P, there is no danger of the mental cause being "excluded" at all. In fact, the rejection of *closure* comprises a rather neat, straightforward response to the causal exclusion argument. The dualist need not defend the possibility of simultaneous causation, nor of efficacious yet systematically overdetermined mental causes, nor need she appeal to *fact* causation in order to understand the reality of mental causation. She can simply posit the existence of irreducibly mental causes of physical events and leave it at that. Without *closure*, the causal exclusion argument never gets off the ground. (This is, of course, why some have dubbed it the "causal closure argument.")[158]

If, by rejecting *closure*, the dualist can avoid the difficulties of the causal exclusion argument *and* affirm the possibility of libertarian agency, then why don't all dualists reject it? What is the evidence in favor of *closure*? What, if anything, ought to prevent the dualist from rejecting this, admittedly widely-held, principle?

[158] See, for example, E.J. Lowe's "Causal Closure Principles and Emergentism" *Philosophy,* 75, 571-585 (2000).

§2.1 On Scientific Respectability

To the question, "Why don't all dualists reject causal closure?" there is one surprisingly simple answer: they don't want to be accused of being anti-scientific, ignorant dogmatists. (There is, for reasons we will consider shortly, a commonly-held belief that the rejection of causal closure is the rejection of science (with a capital "S"), and that no self-respecting philosopher of mind ought to adopt such a position.[159]) More charitably, some substance dualists may choose not to reject causal closure because they recognize a near-consensus and would prefer, if at all possible, to avoid having to go against this consensus. This need not be for cowardly reasons, nor for intellectually lazy ones, but rather for quite respectable ones: very many intelligent people have thought about the question of causal closure, and most of them have come to affirm its truth. Barring reasons to reject causal closure, the dualist might be well advised, *commended* even, for conceding to the majority opinion on this question.

What I hope to have shown, however, is that there *are* good reasons for rejecting *closure*. Indeed, there are *excellent* reasons, and the dualist who is on the fence, so to speak, ought now to jump *off* the fence. As I argued in Chapter Three, the interactionist dualist who accepts *closure* will likely *also* have to accept the impossibility of libertarian freedom. In contrast, the *rejection* of *closure* enables the dualist to respond to the causal exclusion argument, and to do so in an exceedingly straightforward manner. For these

[159] For one example, see David Papineau's *Thinking About Consciousness*. Although he comes to see the difficulties with assuming completeness, or closure, he notes the following: "The one assumption I did expect to be uncontroversial was the completeness of physics. To my surprise, I discovered that a number of my philosophical colleagues didn't agree…My first reaction to this suggestion was that it betrayed an *insufficient understanding of modern physics*." (p.45, my emphasis) (Oxford: Oxford University Press, 2002)

reasons, unless the evidence in favor of causal closure is so strong as to outweigh both of these considerations, the interactionist dualist ought simply to reject *closure*.[160]

§2.2 Evidence of Causal Closure: From Completeness to Closure

What, then, is the force behind this consensus? What is the *evidence* in favor of causal closure? According to many physicalists, the evidence for closure is the success—past, present, and future—of science. In Chapter Three, we discussed the close connection between the causal closure of the physical domain and the completeness of physics.[161] We noted that the two seem to go hand in hand, such that a complete physics should indicate a causally closed physical domain, and a causally closed physical domain should evidence a complete physics. In light of this connection, many feel that a denial of causal closure is tantamount to a dismissal of physics as a complete, or completeable, enterprise.

Jaegwon Kim, for example, writes that, if causal closure were false, then "complete physics would be impossible, even as an idealized goal." After all, if the physical domain is not causally closed, then the physical domain, alone, will not suffice for a complete causal history of all physical events. Instead, for some physical events, we will have to "go outside the physical realm and appeal to nonphysical causal agents and

[160] For reasons that will become clear, I don't actually want to suggest that the dualist reject *closure* as *false*. Instead, the dualist ought to reject *closure* as inadequate support for a causal exclusion argument. It *might* be false, but—depending upon how one defines "physical"—it might very well be *true but nonthreatening*. I will discuss this distinction in §3.2 and, in greater detail, in §6.

[161] See Chapter Three, §4.2.

laws governing their behavior!"[162] Even if it is *possible*—as Kim would surely concede—that physics will one day turn out to be incompletable, it is another thing entirely to conclude incompletability *now*. As such, claims the physicalist, we ought not to affirm a doctrine that commits us to the inevitable failure of physics, and the denial of causal closure does just that.

Similarly, as we saw in Chapter Three, Andrew Melnyk argues that the success of physics today can serve as *positive* evidence of causal closure.[163] In "Some Evidence for Physicalism," he writes,

> Although the claim that the physical is causally closed is not explicitly stated in physics textbooks, it may nonetheless be inferred from claims that *are* explicitly stated in physics textbooks. According to the textbooks, then, contemporary physics has succeeded in finding sufficient physical causes for physical effects of very many kinds; and it has found no physical effects at all for which it is necessary…to invoke non-physical causes. But current physics' success to date in finding that *many* physical events have sufficient physical causes provides inductive evidence that *all* physical events, including *both* unexamined physical events *and* examined-but-as-of-yet-unexplained physical events, have sufficient physical causes.[164]

According to Melnyk, the success of physics is so great as to ground a positive, inferential argument for causal closure. Not only should we hold out hope for the future of physics, but we should conclude—here and now—that causal closure is true, and physics completable.

The following two premises are central to Melnyk's argument:

(1) Current physics has succeeded in finding sufficient physical causes for physical effects of many kinds.

---

[162] Jaegwon Kim, *Philosophy of Mind* (Boulder: Westview Press, 1996, p.147) Cited in Barbara Montero, "Varieties of Causal Closure" in Sven Walter & Heinz-Dieter Heckmann *Physicalism and Mental Causation* (Exeter: Imprint Academic, 2003) p.179

[163] Chapter Three, §4.2

[164] In (Walter & Heckmann, 2003) p160-161

(2) Current physics has found no physical effects at all for which it is necessary to invoke nonphysical causes.

From these premises, we are to conclude causal closure, which Melnyk defines here as:

(CC) All physical events, including both unexamined physical events and examined-but-as-of-yet-unexplained physical events, have sufficient physical causes.

From the fact that physics has found *many* sufficient causes, and *no discernible gaps*, we are to infer causal closure.

What should we make of this argument? There are, I think, a number of ways that a dualist might respond here. First, she might note the difficulties raised by quantum indeterminacy—difficulties that, in a concessionary footnote, Melnyk himself acknowledges.[165] Alternatively, she might question the *structure* of the argument, which seems to move from the claim "We haven't found any nonphysical causes" to the conclusion "There *aren't* any nonphysical causes."[166] In what follows, however, I will not pursue either of these worries. Instead, I will focus on an assumption that underlies *both* Melnyk's inductive argument and Kim's claim that the completeness of physics depends

[165] Melnyk writes, "I should point out that the formulation of the closure principle in the text is not quite right, since it speaks of 'sufficient' physical causes of physical effects, whereas, given the indeterminism of quantum mechanics, no physical events have sufficient physical causes. To avoid this difficulty, we should instead express the closure principle as the claim that the chances of all physical events are determined by earlier physical events plus physical laws, including the irreducibly statistical laws of quantum mechanics. I ignore this refinement in the ensuing discussion."(Melnyk 2003, p.160, fn7)

Melnyk himself is not troubled by this, and suggests that things can simply be rewritten in terms of "chance-fixing." I will not address this worry here, but for an argument against the claim that chance-fixing works just as well as sufficient causation, see E.J. Lowe's "Physical Causal Closure and the Invisibility of Mental Causation." (in Sven Walter & Heinz-Dieter Heckmann, *Physicalism and Mental Causation* (Exeter: Imprint Academic, 2003) pp.137-154. (See especially pp.143-145.)

[166] For more on this worry, see Barbara Montero's "Varieties of Causal Closure" in (Walter & Heckmann 2003), pp. 173-190. (See especially pp.184-185.)

upon the causal closure of the physical domain. I hope to show that this assumption, though widespread, is unfounded.

§2.3 Closure and Completeness Revisited

In formulating his inductive argument for physicalism, Melnyk appeals—implicitly, but crucially—to the following conditional:

(CF) If physics is complete, then the physical world is causally closed.

To see why it is that Melnyk's argument rests upon (CF), note that, if (CF) were false, the success of physics could not serve as evidence of the closure of the physical world. According to Melnyk, the likely completability of physics—as evidenced by its explanatory and predictive success—is evidence of the closure of the physical world. Unless the former entailed the latter, it's hard to see why evidence of the former would serve as evidence of the latter.

Similarly, recall Jaegwon Kim's worry that, if causal closure were false, then "complete physics would be impossible, even as an idealized goal." In making this claim, Kim affirms the contrapositive of (CF):

(CP) If the physical world is *not* causally closed, then physics is *not* complete.

In both cases, the message is clear: the completeness of physics and the causal closure of the physical domain rise and fall together; the one cannot be had without the other.

If this is true, then it must be the case that there is a tight connection between physics and the physical world, such that the latter can be defined in relation to the

former.[167] To see why, suppose that physics and the physical were *not* so related, and the definition of "physical" made no appeal to physics. On this supposition, an entity could be the proper study of physics without, thereby, being physical.[168] If the domain of physics included nonphysical entities, however, then physics might be complete only insofar as it made appeal to physical *and* nonphysical entities. There would, on this account, be *no* reason to believe that the completeness of physics entailed, or even *supported*, the causal closure of the physical domain—for completeness of this sort is wholly compatible with there being nonphysical singular causes of physical events. Indeed, if "physical" does not mean, roughly, "an object of physics," then physics might be complete and the physical world might be really rather *far* from being causally closed.

The completeness of physics can only support the closure of the physical world if the entities in virtue of which physics is complete are the same entities that are said to constitute a causally closed domain. Absent this connection, completeness and closure come apart. The trouble, as we are about to see, is that this tight connection is difficult, if not impossible, to affirm.

167 Barbara Montero offers a more detailed defense of this claim in (Montero 2003, pp178-179.)

168 That is to say, it is theoretically possible that this be the case. Perhaps there is a definition of "physical" that makes no *explicit* reference to physics but which, nevertheless, ensures covariance between physical entities and the domain of Physics. I know of no such definition.

§3.1 Hempel's Dilemma

Hempel's Dilemma takes its name from Carl Hempel, who first posed the problem in his 1969 "Reduction: Ontological and Linguistic Facets."[169] It is, primarily, a dilemma for those who would affirm the truth of physicalism—the claim, roughly, that everything that exists is physical. However, as we will see, it is not *only* a problem for the physicalist. Instead, anybody who affirms the causal closure of the physical world must take a position with respect to the challenge of Hempel's Dilemma.

The problem, Hempel notes, is that physics is not presently a complete science. For that reason, the physicalist who defines "physical" in terms of physics will have to make clear *which* physics she means to refer to: present-day physics, or some future or idealized physics. Should she choose to define "physical" in terms of current physics, and do define "physicalism" accordingly, she will be left with a theory that is very likely false. After all, if "physical" means "an object of present-day physics," then physicalism amounts to the claim that everything that exists is an object of present-day physics. If this is true, then the discovery of new physical entities is impossible; anything that the physicists may find tomorrow, or in ten years time, *cannot be counted as physical*. Surely the physicalist does not mean to affirm that the physicists have *already discovered* all that there is. Yet, should the physicalist choose the first horn of this dilemma and define "physicalism" in terms of present-day physics, that is precisely what her view would amount to. On the first horn of Hempel's Dilemma, then, the truth of physicalism is extraordinarily unlikely.

---

[169] Hempel, C. (1969): 'Reduction: Ontological and Linguistic Facets', in Patrick Suppes, Sidney Morgenbesser and Morgan White (eds.), *Philosophy, Science, and Method: Essays In Honor of Ernest Nagel* (pp. 179–199), St. Martin's.

The second horn of the dilemma is best understood as two related worries, both stemming from a single concern. Broadly understood, the problem with defining physicalism in terms of *future* physics is that we don't know what future physics will look like. Andrew Melnyk describes the problem as follows:

> A physicalism whose content was not determinable by us would presumably be impossible for us to support empirically, and might, for all we know, not even exclude from existence the sort of paradigmatically nonphysical items—for example, souls, entelechies, ghosts—which physicalists have traditionally refused to countenance. [170]

There are, as I said, *two* worries here: what we will call the "no content worry" and the "inappropriate extension worry."[171] The "no content worry" is just what it sounds like: if physicalism is defined in terms of a future-based physics, then it's not clear that we will be able to determine the content of the resulting ontology. Until we know what future physics will posit, we cannot know what it means to say that reality is exhausted by the objects of this future physics.

This is especially important for those who, like Melnyk, wish to use the success of *present-day* physics to support the truth of physicalism. For all we know, future physics will look quite different than the science that presently bears the name. For that reason, any inference from the success of *today's* physics to the exhaustive nature of some *future* physics will lack justification. The physicalist who defines physicalism in terms of a future (or idealized) physics runs the risk of losing whatever empirical foundation present-day physics might have provided.[172]

---

[170] Andrew Melnyk, "How to Keep the 'Physical' in 'Physicalism'" *The Journal of Philosophy*, Vol. 94, No. 12 (Dec., 1997), pp. 622-637; p. 622

[171] Jessica Wilson, "On Characterizing the Physical." (*Philosophical Studies* (2006) 131:61–99) p.68

[172] It is, I hope, clear that the charge raised against a future-based physicalism applies with at least equal force to a physicalism based on an *idealized* physics. If we lack knowledge about what physics will say in 25 years, we *certainly* lack knowledge about what it would say were to it reach its ideal end.

The second worry is this: because we don't know what the physicists will eventually find, it's very difficult to say what they will *not* find. That is, it is difficult—if not impossible—to rule out the eventual discovery of things that would, intuitively, make physicalism false. (This is why Jessica Wilson has dubbed this the "inappropriate extension" worry—for such a physicalism might include in its ontology things that seem not to belong in a physicalist ontology.) What is to prevent the physicists from discovering *sui generis* mental forces, for example, or psychic or protopsychic mental laws, like the ones Chalmers envisions? This concern, the "inappropriate extension worry," constitutes the second problem that arises when one defines "physical" in terms of future physics. Even if the first worry could be met, and a future-based account of "physicalism" could be said to have an adequate degree of content, it might nevertheless fail to be the right *sort* of content. It's hard to see how a "physicalism" that allows for Cartesian souls, for example, would be a physicalism at all.

We have, then, the two horns of Hempel's Dilemma. Should the physicalist take the first horn, and define her ontology in terms of present-day physics, then the resulting physicalism will be very likely false. Should she instead take the second-horn, and define physicalism in terms of some future, or idealized, physics, then the resulting construal of physicalism will run the twofold risk of (a) lacking content completely, and of (b) allowing for the possibility of "physical" ghosts.

§3.2 The Dilemma Applied

In the introduction to this chapter, I claimed that the dualist ought *not to worry* about accepting causal closure. We are now in a position to see what I meant by that, and why it's so significant. Any claim as to the causal closure of the physical world must, of course, invoke the term "physical." In light of Hempel's Dilemma, we may—and, indeed, should—ask: *what is it* that constitutes a causally closed domain? Is it the domain constituted by the objects of *present* day physics, or of some *future*, or idealized physics? The truth of causal closure cannot accurately be assessed until we have disambiguated these two interpretations of the principle. For that reason, a closure principle that has *not* been so disambiguated is not something that *anyone*—dualist or otherwise—ought to grant, for it's not at all clear, in doing so, what is being granted! Ultimately, as I will argue in §6, *neither* will suffice for a (convincing) causal exclusion argument against dualism. If the first is affirmed, then causal closure is false; if the second is affirmed, it is compatible with fundamentally mental causation. In both cases, the result is not something that a dualist—even an interactionist substance dualist—will need to worry about.

To see why this is the case, note that both horns of Hempel's Dilemma apply directly to the question of closure. Should the advocate of closure grasp the first horn, and define "physical" in terms of present-day physics, then she will be left with a causal closure principle that is (almost) certainly false. Consider, once again, the closure statements proposed by Kim and Melnyk, respectively:

> *Closure*: If a physical event has a [sufficient] cause that occurs at *t*, it has a physical [sufficient] cause that occurs at *t*. [173]
>
> (CC) All physical events, including both unexamined physical events and examined-but-as-of-yet-unexplained physical events, have sufficient physical causes.

A moments reflection will tell us why, on the first horn of Hempel's Dilemma, *neither* principle is likely to be true: it is extraordinarily unlikely that all of the causally relevant features of the world have already been discovered by the physicists. The physicists do not claim that physics is complete, and I see no reason at all why we should think that it is. Yet, if "physical" means (roughly) "an object of present day physics," then a *presently complete* physics is precisely what *closure* and *CC* entail. Likewise, if present day physics is *not* complete, if instead there are at least some physical events the causes of which the physicists have *not* yet discovered, then both *closure* and *CC* (so construed) are false.

What about the second horn? There are, as we have seen, two aspects of the second horn of Hempel's Dilemma: the *no content worry* and the *inappropriate extension worry*. For now, I will mostly bracket the first of these concerns. I do this for two reasons: First, I have been persuaded that the *no content worry* is not much of a worry after all. As we shall see, there are ways of defining a future-based physics without sacrificing content. Second, if I am wrong about this and a future-based account of the physical is an empty account, then a future-based causal closure principle will be empty as well. One needs no argument to see why an empty concept is a nonthreatening one.

[173] Jaegwon Kim *Physicalism or Something Near Enough.* (Princeton: Princeton University Press, 2005) p.43.

The trouble with a future-based causal closure principle, then, is the problem of inappropriate extension. Once more, our closure principles:

> *Closure*: If a physical event has a [sufficient] cause that occurs at *t*, it has a physical [sufficient] cause that occurs at *t*. [174]
>
> (CC) All physical events, including both unexamined physical events and examined-but-as-of-yet-unexplained physical events, have sufficient physical causes.

Now suppose that fundamental mentality exists, and is causally active in the world—and *not* merely with respect to other fundamentally mental entities. If this is true, then an idealized physics—a physics that is *complete*—must take these entities into account. A future-based physics, therefore, *may very well* take these into account. All of this may seem to presuppose the existence of fundamental mentality, but notice that the mere *possibility* of fundamental mentality suffices for the *possibility* of such entities being accounted for by physics in the future.

If it is possible that fundamental mentality exists, then—on the second horn of the dilemma—it is possible that there are fundamentally mental *physical* entities.[175] In §6.3, we will consider in greater detail what it would mean for dualism if the mental were incorporated into the physical in this way. For now, it will suffice to note that such an incorporation (a) is wholly compatible with both *closure* and *CC*, understood in terms of the second horn, and (b) undermines the argumentative force of the causal exclusion

[174] (Kim 2005, p.43)

[175] If this sounds like an oxymoron, we ought to note that it is not. The definition of "physical" that we are presently considering does not preclude the possibility of fundamental mentality, and "mental" need not mean "nonphysical."

argument.[176] We will return to both of these points throughout the remainder of this chapter, as we consider various physicalist responses to Hempel's Dilemma.

To summarize, Hempel's Dilemma reaches beyond the physicalist's need to define *physicalism*, and applies with equal force to the challenge of defining causal closure. The physicalist who adopts the first horn of the dilemma must address the charge that the resulting physicalism, and closure principle, are very likely false. The physicalist who adopts the second horn must likewise contend with the charge that her physicalism and closure principle, *even if true*, are compatible with something that looks an awful lot like interactionist substance dualism.[177]

§4.1 The First Horn: Andrew Melnyk's Physicalism

In "How to Keep the Physical in Physicalism," Andrew Melnyk presents what is, as far as I know, the only attempt at grasping the first horn of Hempel's Dilemma.[178] Melnyk grants that defining physicalism in terms of present-day physics has the unfortunate result of rendering physicalism (very probably) false. Nevertheless, he advocates doing just that. Where Melnyk's argument gets interesting is in his *defense* of this position. After all, shouldn't a physicalist who believes that physicalism is very likely

[176] Briefly, the compatibility of a future-based closure principle with causally active, fundamentally mental entities undermines the causal exclusion argument by falsifying premise (3)—the claim that the mental cause is not identical with the physical cause that closure demands. We will discuss this in greater detail towards the end of this chapter.

[177] Note that it is *not*, strictly speaking, compatible with substance dualism itself—for dualism requires a duality of substance, and our present proposal would involve a *unified* substance that happens to include fundamental mentality.

[178] *The Journal of Philosophy*, Vol. 94, No. 12 (Dec., 1997), pp. 622-637

false find another position? Melnyk says that he should not, and that, on the contrary, the physicalist need not be at all troubled by the likely falsehood of physicalism.

Melnyk summarizes the argument against taking the first horn, roughly, as follows: current physics is very probably incomplete, and so a current-physics based physicalism is very probably false. If *this* is what is meant by physicalism, then we ought not to be physicalists, for physicalists are committed to the (at least likely) *truth* of physicalism. In response, he writes the following:

> My reply to this argument is to challenge its final step, that is, the inference that a physicalist should abandon physicalism just because physicalism is very likely false. The argument assumes that a physicalist is someone who must assign a high, or even very high, probability to the thesis of physicalism…But I deny this assumption, claiming that a physicalist need *not* assign a high probability to physicalism, and can therefore comfortably live with the result that physicalism has a very low probability.[179]

Now, on a first reading, this sounds kind of crazy. After all, what *is* a physicalist if not a person who believes physicalism to be true?

In defense of the claim that a physicalist need not be committed to the truth of physicalism, Melnyk offers what I will call the *SR Argument*. Central to the SR argument are the following definitions:[180]

> (SR) To take the SR attitude toward a hypothesis is (1) to regard the hypothesis as true or false in virtue of the way the mind-independent world is, and (2) to assign the hypothesis a higher probability than that of its *relevant rivals*.
>
> (RR) Hypothesis H1 is a relevant rival to H2 if and only if (a) H1 is sensibly intended to achieve a significant number of H2's theoretical goals; (b) the hypotheses, H1 and H2, fail to supervene on one another; and (c) H1 has actually been formulated.

[179](Melnyk 1997, p. 624)

[180] (Melnyk 1997, p. 625-626)

According to Melnyk, the SR attitude is "the attitude that those who have broadly scientific realist and antirelativist intuitions take toward what they regard as the best of current scientific hypotheses." (626) An advocate of, say, string theory may not necessarily believe that the theory is *true*, at least not in its entirety. Instead, she might simply believe that string theory is *better than the available alternatives*. The fact that she is not firmly convinced of the *truth* of string theory is not, says Melnky, reason enough to deny that she is a string theorist.

In the same way, argues Melnyk, a physicalist need only adopt the SR attitude towards physicalism in order to affirm physicalism. As long as she believes that physicalism is:

> (a) true or false in virtue of the way the mind-independent world is, and
>
> (b) more likely to be true than any other *genuinely distinct, actually formulated* ontology

she may call herself a physicalist. Furthermore, not only is it possible to adopt the SR attitude towards a theory without affirming the *truth* of that theory, one need not even believe that said theory is *at all likely* to be true. After all, if none of the available theories are likely to be true, then the *most* likely need not be very likely at all. Melnyk concludes, "Therefore, to be a physicalist does not require regarding physicalism as likely to be true (let alone very likely to be true)."[181]

To summarize, then, Melnyk suggests that the physicalist define "physical" with reference to current physics. Accordingly, the only things that count as physical are those things—entities, forces, laws—with which the physicists are already acquainted. Furthermore, Melnyk offers the following definition of physicalism:

---

[181] (Melnyk 1997, p.625)

> Physicalism is roughly the thesis (1) that every entity is either itself a physical entity or is exhaustively composed, ultimately, of physical entities, and (2) that every property is either itself a physical property or is realized, ultimately, by physical properties.[182]

Given the very likely discovery of new entities, Melnyk's physicalism is highly unlikely to turn-out to be true. Still, claims Melnyk, the likely falsehood of physicalism need not dissuade the physicalist from affirming it. Instead, as long as she is able to adopt the SR attitude towards physicalism, she ought not to worry about the (probably inevitable) ultimate falsification of her theory of choice.

§4.2 The First Horn: Relevant Rivals

What should we make of this argument? As I see it, there are two questions that we ought to ask in response to the SR argument: (1) Is it enough that the physicalist adopt the SR attitude towards physicalism? (2) Is the SR attitude something that a physicalist can justifiably adopt towards a theory that is, admittedly, probably false? In answer to the first question, I am inclined to think that, yes, a person who is able to adopt the SR attitude towards physicalism can reasonably be called a physicalist. The second question is a bit more complicated. There, I claim that a physicalist *cannot* justifiably adopt the SR attitude towards Melnyk's Physicalism—not because it is probably false, but rather because of the *reasons* for which it is probably false. To see why, it will be helpful to look more carefully at what Melnyk means by a "relevant rival."

Suppose I were to offer the following rival hypothesis to physicalism:

[182] (Melnky 1997, p. 622)

> *Antiphysicalism*: (1) not every entity is either itself a physical entity or exhaustively composed, ultimately, of physical entities and, (2) not every property is either itself a physical property or realized, ultimately, by, physical properties.

If antiphysicalism qualifies as a rival hypothesis to physicalism, then no physicalist can rationally adopt the SR attitude towards physicalism. The falsity of physicalism entails the truth of antiphysicalism; accordingly, even if the probability of physicalism being true were as high as .49, antiphysicalism would remain the more probable hypothesis. Furthermore, given the definition of physicalism in terms of present day physics, no physicalist believes the probability of physicalism to be anywhere near as high as .49. If antiphysicalism is a viable alternative, then, it follows that no physicalist can adopt the SR attitude towards physicalism.

Melnyk anticipates this line of reasoning. In fact, it is for this reason that he includes a detailed account of what does, and does not, qualify as a rival hypothesis. Specifically, he writes:

> One especially important consequence of (RR) is that the sheer negation of a hypothesis, unsupplemented by any other claims, does *not* count as a relevant rival to the hypothesis, since the unsupplemented negation of a hypothesis…cannot sensibly be intended to achieve the theoretical goals of the hypothesis.[183]

Instead, in order for the negation of a hypothesis to count as its rival, it must be supplemented with some additional claims—claims that offer an alternative approach to the theoretical goals of the original hypothesis. For example, Melnyk notes that while "atheism unadorned" does not qualify as a rival of theism, atheism conjoined with "the findings of contemporary science" *does*.[184] Presumably, this is because the simple denial of the existence of God cannot explain the phenomena that theists attribute to God. The

[183] (Melnyk 1997, p.627)

[184] (Melnyk 1997, p.627)

additional components, the findings of science, *can* go some distance towards accounting for those phenomena, and so the conjunction of the two meets criterion (a) of RR. Antiphysicalism alone will not, therefore, qualify as a relevant rival to physicalism.

What *will* count? Consider again Melnyk's definition of a relevant rival:

> (RR) Hypothesis H1 is a relevant rival to H2 if and only if (a) H1 is sensibly intended to achieve a significant number of H2's theoretical goals; (b) the hypotheses, H1 and H2, fail to supervene on one another; and (c) H1 has actually been formulated. (626)

It is in virtue of the first criterion, (a), that antiphysicalism fails to qualify as a relevant rival to physicalism. The latter two criteria serve primarily to ensure that the relevant rival is, in fact, a *rival*: (b) tells us that the rival hypotheses are logically distinct from one another, and (c) prevents us from appealing to some future, presently unarticulated theory. In contrast, the first criterion is crucial to determining the *relevance* of a potential "relevant rival."

What, then, are the "theoretical goals" of Melnyk's physicalism? In order better to understand what a relevant rival of his own theory would be, Melnyk suggests that we understand "physicalism" as the conjunction of the following two theses:

> (1) There is some science, *S*, distinct from the totality of all the sciences, such that every entity (property) is either itself mentioned as such in the laws and theories of *S* or is ultimately constituted (realized) by entities (properties) mentioned as such in the laws and theories of *S*.
>
> (2) *S* is current physics. (633)

So understood, the theoretical goals of physicalism are to assert (1) the existence of one fundamental, ontologically exhaustive science and (2) that current physics *is* that science. More broadly understood, the theoretical goal of Melnyk's physicalism is to explain the relationship between current physics and the world.

In light of these goals, Melnyk notes that the potential relevant rivals to (his) physicalism will fall into the following two categories: those that affirm (1) but deny (2), and those that deny (1). In what follows, I will not consider relevant rivals of the second kind.[185] Instead, I will devote the remainder of this chapter to rivals of the first kind—those that affirm the existence of a fundamental, ontologically exhaustive science, but deny that current physics is that science. My reasons for this are simple: I believe that there are relevant rivals of this variety that a committed physicalist really ought to deem more likely to be true than Melnyk's physicalism. If I am correct, then the SR attitude is not a justifiable attitude to adopt towards Melnyk's physicalism.

Before considering these rival theories, however, I want to note what Melnyk says about hypotheses of this variety. He first considers the possibility of choosing some *other* physical science—specifically, he suggests biology—as the fundamental, ontologically exhaustive science posited in premise (1). Because physics is more fundamental than biology, and because biology *also* has a history of changing and developing, Melnyk notes that a relevant rival that replaced physics with biology would be *less* likely to be true than physicalism.

Of course, nobody *claims* that biology can play the role that physicalism ascribes to physics. Instead, the more common rival to physicalism, according to Melnyk, is dualism. He writes:

> The best-known relevant rival [that affirms (1)] is traditional dualism, which I interpret as the view that, to put it very crudely, physicalism is true of everything except the mind: there is a basic science, but it is the *conjunction* of physics and folk psychology…

[185] For Melnyk's discussion on why these rivals will not be more probably than his physicalism, see (Melnyk, 1997, p. 634-635.)

If Melnyk is correct, then dualism—which he takes to be a supplemented physicalism of sorts—is the most common alternative attempt at achieving the theoretical goals of physicalism. Where physicalism posits a certain relationship between physics and the world, dualism posits a *similar* relationship, but supplements the role of physics with that of folk-psychology. Dualism is, therefore, a relevant rival to physicalism.

Yet, so understood, dualism cannot be more likely to be true than physicalism is. After all, as Melnyk notes, to whatever extent the inevitable progress of physics makes physicalism likely to be false, it will, to that same extent, make dualism likely to be false.[186] If Melnyk's physicalism makes up a large part of dualism, then dualism will be equally endangered by the progress of physics. For that reason, Melnyk tells us, physicalism is *more likely to be true* than dualism, its "best known relevant rival."[187]

§4.3 The First Horn: A More Relevant Rival

We are, at last, in a position to see where Melnyk's argument goes wrong. We began with the question of how, in light of Hempel's Dilemma, one ought to define the physical. Contrary to most physicalists, Melnyk suggests that we grasp the *first* horn of the dilemma and define physicalism with respect to current physics. Noting the likely falsehood of physicalism so construed, Melnyk argues that a physicalist need only affirm

[186] There is one way that this would not be true: If the physicists were to find things that validate the claims of folk-psychology with respect to the mind, then dualism would fare better than physicalism in the face of such findings. Still, as a general point, the fact remains: a dualism that includes most of the claims of physicalism is not likely to be more probable than physicalism is.

[187] Again, the claim is only that dualism is the best known relevant rival *of this variety*. Nevertheless, because Melnyk takes this approach to be the stronger of the two, the claim here is intended to be a fairly strong one.

the *greater likelihood* of physicalism with respect to its relevant rivals. He then attempts to show, albeit briefly, that the most likely candidates for a relevant rival to physicalism are less likely than physicalism is, and so a physicalist can justifiably affirm physicalism.

The trouble is, Melnyk's physicalist doesn't just have to worry about rivals to physicalism *in general*, she has to worry about rivals to physicalism *defined in terms of current physics*. In light of Hempel's Dilemma, isn't it clear that *other formulations of physicalism* are the most obvious candidates for a relevant rival for *this* construal of physicalism? After all, the whole point of Hempel's Dilemma was to show that there are a number of things that might be meant by "physicalism," and the truth—or *likelihood*—of the position cannot be assessed until these ambiguities have been sorted out. Having disambiguated things in one way, furthermore, does not exempt the physicalist from keeping the alternative possibilities in mind.

To put the problem in Melnyk's terms, recall that he divided possible relevant rivals into two categories—those that affirm only the first of the following two theses, and those that deny even the first:

> (1) There is some science, *S*, distinct from the totality of all the sciences, such that every entity (property) is either itself mentioned as such in the laws and theories of *S* or is ultimately constituted (realized) by entities (properties) mentioned as such in the laws and theories of *S*.
>
> (2) *S* is current physics. (633)

Suppose, then, that I accept (1) but reject (2). I might do this for one of two reasons: I might believe that *S* is some science wholly unrelated to physics, or I might believe that *S* is physics, but *not current physics*. Indeed, the (many) physicalists who advocate a future-based definition of the physical are, it seems, doing just this.[188] Such physicalists

[188] See, for example, J. Poland *Physicalism: the Philosophical Foundations*, (Oxford: Clarendon Press, 1994), (Wilson 2006) and (Dowell 2006)

affirm the existence of a fundamental, ontologically exhaustive science, and they reject the claim that *current* physics is that science—but they offer in its place a future, or idealized, version of physics. Surely the accounts of this variety are the *most relevant* rivals to Melnyk's physicalism.[189]

More importantly, a physicalism defined in terms of future, or idealized, physics will *of course* be more likely to be true than one based upon current physics. After all, it is the likely *progress* of physics, not the likely future failings, that render a present-based physicalism so very improbable. If, as Melnyk surely believes, the physicists are likely to discover new entities or laws in the future, then a physicalist must concede that future-based physics is *closer* to being ontologically exhaustive than current physics is. That is, she must conclude that a physicalist account that takes the second horn is *more likely to be true* than Melnyk's account. If this is correct, then a physicalist ought not to adopt the SR attitude towards Melnyk's physicalism.

All of this, of course, assumes that a workable future-based physicalism is available. More specifically, these considerations assume that *at least* the "no content worry" can be met. A physicalism that lacks content is neither likely nor unlikely to be true; a meaningful assessment of the likelihood of a philosophical position requires a degree of content that is, at least, sufficient to determine what it would take for the position to be *false*.[190] In what follows, we will consider two alternative physicalist accounts, both of which proceed by grasping the second horn of Hempel's Dilemma.

---

189 Jessica Wilson makes the same point in (Wilson, 2006 p. 66-67.) (I came to this conclusion independently, and only later found that she had done so as well.)

190 I suppose some empty claims are overwhelmingly likely to be true, but not in a meaningful way. Physicalism , for example, is true but tautological if understood as the claim that "everything that exists exists."

Both, I maintain, can meet the demands of the *no content worry*, though only one of the two can also address the *inappropriate extension worry*.

Because there are ways of understanding the physical that do *not* commit the physicalist to the likely falsehood of physicalism, but instead stand to *benefit* from the future developments of physics, I conclude that Melnyk's physicalism ought not to be endorsed by any physicalist. Instead, absent some better approach to the first horn, a physicalist really ought to grasp the *second* horn of Hempel's Dilemma. A causal closure principle stated in terms of the second horn, however, will be inadequate for the role that it is thought to play in the causal exclusion argument. As I will show §6, a future-based understanding of physics entails a causal closure principle that is either *unacceptably stipulative*, or *compatible with interaction*.

§5.1 The Second Horn

The two aspects of the second horn of Hempel's Dilemma are, again, the *no content worry* and the *inappropriate extension worry*. In "The Physical: Empirical, Not Metaphysical," J.L. Dowell gives the following colorful illustration of the two worries:

> Who knows what future people we'll call 'physicists' will study? Given that we have no idea what will be a posit of that theory, we also have no idea what won't. And given that we have no idea what won't be a posit of the theory ultimately developed by physicists, we're unable to identify what would count as falsifying physicalism on the resulting formulation.
>
> To sharpen the objection, suppose that future physicists, perhaps in a series of tragic lab accidents, will go off their collective rockers and take to channeling the dead. This possible scenario highlights just how unconstrained the notion of

'whatever future people we'll call 'physicists' will study' really is.[191]

If, by "future-physics," we mean "whatever enterprise future bearers of the term 'physicists' are engaged in," then a future-physics based physicalism will fail to meet either of the two worries that constitute the second horn of Hempel's Dilemma. On the one hand, an account of this variety would be, from our vantage, unfalsifiable. On the other—and here I admit to stretching Dowell's illustration a bit—this "physicalism" could be true of a world in which the dead are *actually* channeled. A physicalism that is compatible with the possibility of (actual) séances is no physicalism at all.

How *should* a physicalist understand "future-physics?" In two very recent papers, Jessica Wilson and J.L. Dowell assert that there are *theoretical parameters* that establish the limits of what ought to count as "physics."[192] In fact, Wilson and Dowell are largely in agreement as to what these parameters are, and how the future-based physicalist ought to address the *no content worry*. Where they differ, and they do differ, is with respect to the *inappropriate extension worry*. I will treat the two worries in turn.

[191] *Philosophical Studies* (2006) 131:25-60, p.37

[192] (Dowell 2006); Jessica Wilson, "On Characterizing the Physical" *Philosophical Stud*ies (2006) 131:61–99

§5.2 Giving Content to "Future Physics"

According to J.L. Dowell, physics is best understood as an enterprise that meets the following two criteria: [193]

(a) it is a science, and
(b) it treats the most fundamental, or—if there is no *most* fundamental—the *relatively* fundamental, entities.

Dowell suggests four "hallmarks" of a scientific theory. For something to be a science, it ought to have (1) empirically verifiable, explanatory hypotheses, (2) some empirical confirmation of these hypotheses, (3) a unified explanatory account of some empirical generalizations, and (4) additional empirical support—in particular, it should be a "good fit" with established empirical observations.[194] More simply stated, for something to be a science it ought to make *explanatory*, *empirically verifiable* claims about the world, and some of those claims ought to have empirical confirmation.

The second criterion distinguishes physics from, say, biology or chemistry. If there is a most-fundamental level of the actual world's ontology, then an idealized physics will be a science that deals with objects at that level of fundamentality. If, on the

[193] On p.39, for example, Dowell defines a physical theory as "a scientific theory of the world's relatively fundamental elements." Similarly, Wilson defines physics as "a science treating of the relatively fundamental entities." (p.72)

[194] (Dowell 2006, p.39.) Dowell does not claim originality with respect to these hallmarks, but is instead attempting to capture what is already in the philosophy of science literature. For example, she references: Boyd, R. (1983): On the Current Status of the Issue of Scientific Realism ,*Erkenntnis* 17, 135 61-69.

Boyd, R. (1985): 'Lex Orandi est Lex Credendi', in P.M. Churchland and C.A. Hooker (eds.), *Images of Science*, Chicago: University of Chicago Press.

Hempel, C. (1965): 'Aspects of Scientific Explanation', in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: the Free Press.

Hempel, C. (1966): *Philosophy of Science,* Edgewood Cliffs: Prentice-Hall.

other hand, there is *no* most fundamental level, then an idealized physics will be the science of the most fundamental level *relative to that to which we have empirical access*.

If this is what is meant by "physics," then how should we understand physicalism? Dowell writes that "Intuitively, physicalism is the thesis that there's nothing 'over and above' the physical." For a slightly more extensive account, recall Melnyk's definition:[195]

> Physicalism is roughly the thesis (1) that every entity is either itself a physical entity or is exhaustively composed, ultimately, of physical entities, and (2) that every property is either itself a physical property or is realized, ultimately, by physical properties. (622)

A future-based physicalism, then, amounts to the claim that there is nothing "over and above" the relatively fundamental entities; every entity (property) is either itself a relatively fundamental entity, or is exhaustively composed (realized) of relatively fundamental entities.

In this way, Dowell notes, the future-based physicalist can meet the *no content worry*—for, contrary to the objection, it is in fact *clear* what would count as a falsification of *this* physicalism. If there are entities that are *less* fundamental (i.e., more complex) than the (relatively) fundamental entities, and those entities are not in any way reducible to the fundamental ones, then Dowell's physicalism is false. Likewise, if there are relatively complex properties, and those properties are not realized by any relatively

---

[195] Both Dowell and Wilson equate physicalism with the claim that there is nothing "over and above the physical." (Wilson p.62, Dowell p.25.) Unlike Melnyk, they do not explicitly distinguish between the proper objects of physics and those things that are *composed* of the proper objects of physics. Because I think this is a valuable distinction, and because I think both Wilson and Dowell would also affirm Melnyk's definition, I have chosen to use Melnyk's definition here as well.

I do think one proviso is necessary: the world "realized," in this context, should not be read as entailing all that Melnyk means by "realization physicalism." Instead, I mean only to include the possibility of there being higher-level entities or properties that are—*somehow*—composed out of the more fundamental ones.

fundamental properties, then Dowell's physicalism is false. Stated in terms of the science of physics, if it turns out that there are things that cannot be incorporated into physics so understood, then physics is not the ontologically exhaustive science that physicalism says it is, and physicalism is false.

Furthermore, Jessica Wilson—who defines physics, like Dowell, in terms of relative fundamentality—notes that *some* of the content of a future physics will be the content of *present* day physics. After all, current physics is a science that treats the most fundamental entities *relative to the limits of our knowledge*. Of course, current physics is overwhelmingly likely to be false.[196] For that reason, Wilson includes in her definition of a "physical entity" that it be "treated, *approximately accurately*, by current or future…versions of fundamental physics." (72) (emphasis added) Assuming, then, that current physics gives an approximately accurate account of some entities, it follows that those entities form a part of a future-based physicalism as well.

It seems, then, that the *no content worry* can be met. The objects of current physics give us *some* idea of what future physics will look like, (unless, of course, current physics is a *complete* failure), and the possibility of there being irreducible entities of relatively high complexity provides a clear account of how the world would have to be for future-based physicalism to be false. Future-based physicalism, defined in terms of a science that treats the relatively fundamental entities of the world, is hardly an empty theory.

[196] Wilson actually maintains that present physics is *certainly* false, owing to the inconsistency of the conjunction of The Standard Model and General Relativity. (Wilson, 2006 pp.62-65)

§5.3 The First Horn Revisited

Before moving on to the *inappropriate extension worry*, I wish to note the following: If this is correct, and the no content worry can be met, then Melnyk's response to Hempel's Dilemma fails—*even if* the inappropriate extension worry cannot be met. Physicalism defined in terms of future physics clearly meets the criteria for a relevant rival to (Melnyk's) physicalism.

> (RR) Hypothesis H1 is a relevant rival to H2 if and only if (a) H1 is sensibly intended to achieve a significant number of H2's theoretical goals; (b) the hypotheses, H1 and H2, fail to supervene on one another; and (c) H1 has actually been formulated. (626)

If, as Melnyk supposes, substance dualism is intended to achieve a significant number of the theoretical goals of his physicalism, then physicalism based on future physics *surely* is. Like Melnyk, Dowell and Wilson propose accounts that are committed to the existence of a unified, fundamental, ontologically exhaustive science. Criterion (a), then, is met. Further, a physicalism based on future-physics does not supervene on Melnyk's physicalism, for the former might be true in cases where the latter is false—namely, on the assumption that physics progresses. Criteria (b) and, of course, (c) have, therefore, also been met. A physicalism based on future-physics is a relevant rival to Melnyk's physicalism. (Indeed, I would argue that it is the *most* relevant rival.)

In light of this conclusion, it's hard to see how a physicalist could adopt the SR attitude towards Melnyk's physicalism. After all, doing so would require her to assign a higher probability to current-physics based physicalism than to physicalism defined in terms of future-physics. That just can't be right. It simply cannot be the case that physics *today* is more likely to be the fundamental, ontologically exhaustive science than physics

in, say, 50 years is. When you include the possibility of *idealized*, *complete* physics, the matter really is decided. Unless the physicalist is ready to say that the physicists are done, that they have gone as far as they can go, she cannot say that a current-physics based physicalism is more probable than one based on future-physics.[197]

§5.4 The Second Horn: Inappropriate Extension

While Wilson and Dowell agree with respect to the *no content worry*, their accounts diverge when it comes to the question of *inappropriate extension*. According to Dowell, it is not up to the philosophers to determine, *a priori*, what kinds of things the physicists are going to discover. Instead, she writes that "if no actual mental property is among the basic physical ones, as seems overwhelmingly likely, that's a matter to be settled a posteriori."[198] Jessica Wilson, in contrast, argues that a physicalist "need not and should not hand over all authority to physics to determine what is physical."[199] Instead, she argues that a "no-fundamental-mentality" constraint ought to be included in one's definition of "physicalism." In what follows, we will consider (albeit briefly) the strengths and weaknesses of these two responses.

We have already seen what Dowell's physicalism amounts to. Her account requires that the entities and properties of the actual world be subsumable under the domain of future physics, where "physics" refers to the science of the relatively

[197] As I noted in fn. 34, Jessica Wilson makes the same point**.** (Wilson, 2006 p. 66-67.)

[198] (Dowell 2006, p.28)

[199] (Wilson 2006, p.69)

fundamental entities. If there is nothing over and above that which is relatively fundamental, then Dowell's physicalism is true. It is, therefore, easy to see how the *inappropriate extension* objection might be raised against Dowell, for if there are relatively fundamental *mental* entities—or relatively fundamental mental *properties*—then they will count as "physical" on this account. In this way, the *inappropriate extension worry* is a real worry for Dowell's physicalism. Physicalism is traditionally taken to exclude the possibility of irreducible mentality; Dowell's physicalism allows for it.[200]

Nevertheless, we should not conflate the claim that *some* "inappropriate" things might count as physical with the stronger claim that *anything at all* might qualify. After all, as Dowell notes, the theory that posits these entities must remain a *science* if it is to remain *physics*. If, therefore, the "physicists" of the future were to incorporate "miracle-performing angels" into their theoretical arsenal, they would cease to be physicists. Dowell writes,

> A miracle-performing angel is an entity whose acts are by definition incapable in principle of being fit into a pattern of explanation characteristic of scientific theories. So if angels were to figure in our ideal physical theory…it would have to be in some mundane sense of 'angel'. They would have to be angels stripped of their miraculous powers and governed by the same laws everything else is.[201]

We will look more carefully at the final claim in this passage shortly. For now, it will suffice to note that there are limits to what kinds of things can be accepted as physical on

[200] At this point, some might question whether or not this account is enough like traditional physicalism so as to count as physicalism. The scope of this chapter does not allow for a detailed treatment of this question, but I will say this: There is an awful lot that goes into being a physicalist account. If, as Dowell argues and as I will argue, the physicalist must break from the *a posteriori* commitments so central to physicalism in order to exclude the possibility of fundamental mentality, then it is at least an open question which approach is *more physicalistic*—the one that preserves a posteriority or the one that excludes mentality.

[201] (Dowell 2006, p.41)

Dowell's physicalism. Those limits exclude the kinds of things that resist the explanatory and predictive grasp of an empirical science, but they do *not* necessarily exclude mentality. To revisit the quotation invoked above, "if no actual mental property is among the basic physical ones…that's a matter to be settled a posteriori."[202]

For Jessica Wilson, this is not enough. Instead, Wilson maintains that it is the job of the *physicalist*, and not just the *physicist*, to determine what is meant by "physical." More specifically, Wilson supplements her future-physics based definition of the physical with a "no fundamental mentality" constraint. She writes, "An entity existing at a world *w* is physical if and only if:

(i) it is treated, approximately accurately, by current or future (in the limit of inquiry, ideal) versions of fundamental physics at w, and
(ii) it is not fundamentally mental (that is, does not individually either possess or bestow mentality)"[203]

(In order for something to qualify as "fundamental physics" at some world *w*, it must be—as we have noted—the *science* of the *relatively fundamental* at world *w*.) Like Dowell, Wilson includes in her definition of the physical an appeal to future-physics; unlike Dowell, her definition is not *exhausted* by this appeal. Instead, Wilson addresses the *inappropriate extension worry* in her definition of "physical." If something is fundamentally mental, then it is not physical on Wilson's physicalism. Even if it were to become the subject of the science of the relatively fundamental, it would *still* not be physical. In this way, Wilson attempts to ward-off the *inappropriate extension worry*.[204]

---

[202] (Dowell 2006, p.28) I have omitted the phrase "as seems overwhelmingly likely." I do this not to misconstrue Dowell's position, but rather for brevity sake. Because I included the full quotation above, I hope that the sentiment of Dowell's claim is, nevertheless, preserved.

[203] (Wilson 2006, p.72)

[204] For a powerful argument to the effect that Wilson's account does not *successfully* keep mentality out of the physical, see Neal Judisch's "Why 'non-mental' won't work: on Hempel's dilemma and the characterization of the 'physical'" (Philosophical Studies (2008) 140:299–318) Briefly stated, Judisch

Wilson and Dowell do not disagree about what it is for something to be *physics*; they disagree only as to what it is for something to be *physical*. This is a significant distinction, and it is one that we ought to keep in mind. Both Dowell and Wilson want to invoke physics in order to define "physical." Both want specifically to include the findings of *future* physics in this definition. Where they differ, then, is in the extent to which they are willing to throw their lot in with the *actual* future of physics. Dowell's account allows for the possibility that the future of physics will diverge from what she, as a physicalist, expects; she ties her physicalism, nevertheless, to the future of this empirical science. Wilson does not. While she grants that the *physics* of the future might include such undesirables as relatively fundamental mental entities, she draws the line at granting that *physicalism* might include this possibility. For this reason, while Wilson's account addresses the inappropriate extension worry in a way that Dowell's cannot, it does so at a cost. As we shall see, *exclusion by stipulation* is not a very powerful philosophical maneuver.

§6.1 Taking Stock of Closure

In light of the considerations raised by Hempel's Dilemma, we have the following competing accounts of what it is for something to be physical (corresponding to Melnyk, Dowell, and Wilson respectively):

Physical$_M$: An entity or property is physical iff it is a posit of current physics.

---

notes that, while Wilson's NFM rules-out *fundamental* mentality, it does not rule-out the possibility of *fundamental protomentality*, as might be posited by a protopsychic or a neutral monist.

> Physical_D: An entity or property is physical iff it is, or will be, a posit of the future (idealized) science of the relatively fundamental.
>
> Physical_W: An entity or property is physical iff it is, or will be a posit of the future (idealized) science of the relatively fundamental *and* it is not fundamentally mental.[205]

With these definitions in hand, we are finally in a position to address, in detail, the questions that were raised in §3.2—namely, the implications of Hempel's Dilemma for causal closure and the exclusion argument. In what follows, I hope to show that *none* of the foregoing versions of physicalism can sustain a causal exclusion argument of any real force.

Consider, once again, the causal exclusion argument against substance dualism:

(1) Suppose mental event M causes physical event P at *t*. (*for reductio*)
(2) P has a sufficient physical cause at *t* as well, call it P*. (*closure*)
(3) M is not identical with P*. (substance dualism)
(4) P cannot have more than one sufficient cause at *t*—unless this is a case of genuine overdetermination. (*exclusion*)
(5) This is not a case of genuine overdetermination.
(6) Then either P* or M is the cause of P, but not both. ((1)-(5))
(7) P*, not M, is the cause of P.
(8) If M causes P at *t*, then M does not cause P at *t*. ⊗

I have maintained that the dualist ought to focus here attention on the causal closure of the physical world, rather than on *exclusion*. In what follows, I will again take Kim's *closure* as a paradigmatic closure statement. That is, I will assume that what holds for *closure* holds for causal closure, broadly understood. There are, to be sure, alternative ways of stating the causal closure of the physical world.[206] Nevertheless, because it is the

[205] I have not included posits of current physics that are treated "approximately accurately" for the following reason: if they are treated approximately accurately by current physics, then they will, presumably, be treated by idealized physics as well.

[206] David Papineau, for example, gives the following definition: "All physical effects are fully caused by purely *physical* prior histories." He then adds, in a footnote, the following disclaimer: "A stricter version of [this principle] would say that the *chances* of physical effects are always fully fixed by their prior

definition of "*physical"* that is central to the problem under discussion, and because *all* causal closure statements must appeal to the physical, I hope it will be clear that the problems for *closure* are problems not for this particular statement of causal closure, but for the claim that the physical world—whatever that may be—is closed in such a way as to exclude irreducibly mental causes.

§6.2 The Causal Closure of the Physical$_M$

As we now know, we cannot assess the strength of the causal exclusion argument until we have disambiguated the word "physical." Suppose, then, that we do so by appealing to Physical$_{M;}$ that is, suppose that we embrace Melnyk's response to Hempel's Dilemma and define the physical in terms of current physics. What happens to premise (2)? Well, (2) is supported by *closure*:

> *Closure*: If a physical event has a [sufficient] cause that occurs at *t*, it has a physical [sufficient] cause that occurs at *t*.

If *closure* is false, then premise (2) is unfounded. Of course, *closure* must *also* be disambiguated in light of Hempel's Dilemma. By applying Melnyk's Physical$_M$ to *closure,* we get:

> *Closure$_M$*: If an *event that is a posit of current physics* has a sufficient cause that occurs at *t*, it has a sufficient cause that occurs at *t that is a posit of current physics*.

---

physical histories…" *Thinking About Consciousness.* (Oxford: Oxford University Press) p17 Papineau's statements differ both from Kim's and from each other. Nevertheless, *both* of them, like *closure*, appeal to "physical" causes and, as such, will need to be disambiguated in light of Hempel's Dilemma. Again, I don't see how the relatively minor differences in competing closure statements will have any bearing on the difficulties raised by Hempel's Dilemma, and for that reason I will not attempt to demonstrate the effects of the dilemma on a variety of principles.

We have already seen why *closure*$_M$ is false. If it were true, then current physics would be complete. Current physics is *not* complete, so *closure*$_M$ is false. It simply isn't the case that the physicists are *already familiar with* all of the causes of all known (caused) events. After all, note how little it takes for *closure*$_M$ to be false: as long as there is one event that *has* a sufficient cause at some time, but lacks a *known* sufficient cause at that time, *closure*$_M$ fails. Surely there are types of events that are (a) causally efficacious in the (known) physical world and (b) currently unknown to the physicists. If so, then *closure*$_M$ is false.

Given Melnky's concession that physicalism, on the first-horn, is likely false, might he argue that *closure*$_M$ need not be true either? That is, can the physicalist simply affirm the SR attitude towards causal closure and leave it at that? I don't think so, for if, as I have argued, a physicalist ought not to adopt the SR attitude towards *physicalism* defined in terms of current physics, she ought *also* not to adopt it towards *causal closure* so defined. Not only is it the case that *closure*$_M$ is false, it is also just obviously *less likely* than a causal closure principle defined in terms of future physics—or so it seems to me. I will not, however, argue this point any further.

Even if the arguments against Melnyk's physicalism fail, and even if the physicalist can justifiably adopt the SR attitude towards *closure*$_M$, this much should by now be clear: *the dualist should feel no compunction to follow suit*. If a causal closure principle is clearly false, then no dualist should affirm it. A causal closure principle defined in terms of current physic, then, cannot ground a causal exclusion argument. The crucial premise of the argument, that sustained by causal closure, has to be at least *likely* to be true in order to be of any argumentative use. If, by *closure*, the physicalist means

closure$_{M,}$ or CC, then the dualist ought simply to reject premise (2) of the causal exclusion argument and leave it at that.

§6.3 The Causal Closure of the Physical$_D$

Suppose, instead, the physicalist grasps the second horn and adopts Dowell's physicalism, Physicalism$_{D.}$ What follows for *closure* then? By substituting Physical$_D$ for "physical," we get:

> *Closure*$_D$: If an *event that is a posit of the future (idealized) science of the relatively fundamental* has a sufficient cause that occurs at *t*, it has a sufficient cause that occurs at *t that is a posit of the future (idealized) science of the relatively fundamental*.

Unlike *closure*$_M$, *closure*$_D$ is not obviously false. For this principle to be false, it would have to be the case that some relatively fundamental event has a cause that is not, itself, relatively fundamental. (If it *were* relatively fundamental, then presumably an idealized physics would talk about it.) Furthermore, if physics is completable, then *closure*$_D$ is true. In order for it to be the case that the *science* of the relatively fundamental is complete, it must be the case that the *domain* of the relatively fundamental is causally closed.[207] An outright rejection of *closure*$_D$, then, would be a far stronger position for the dualist to adopt than a rejection of *closure*$_{M.}$

Fortunately, the dualist *need* not reject *closure*$_D$ in order to address the exclusion argument. As we have seen, the truth of *closure*$_D$ is compatible with the possibility of

[207] I discuss the entailment between causal closure and completeness in Chapter 3, §4.2

fundamental mental causation. Instead, should the physicalist appeal to $closure_D$ in order to ground the exclusion argument, the dualist ought rather to reject premise (3):

(3) M is not identical with P*.

To see why, remember that, on Dowell's physicalism, it is entirely possible for there to be an event that is *both mental and physical*. When $closure_D$ requires that we posit a physical cause of P, it requires only that we posit a cause of P that is *relatively fundamental*.[208] As long as the dualist is prepared to admit the fundamentality of the mental cause, she can—on this definition of the physical—grant its *physicality* as well. If she does, then the causal exclusion argument goes away. Absent premise (3), the distinctness of the mental and physical cause, the argument does not go through.

There are, I think, two potential objections that must be addressed at this point. First, if the dualist grants that the mental is physical, in what sense does she remain a *dualist*? Second, even if what we now call the mental can be incorporated by the physical, can any of the important *features* of mentality survive? What is to prevent the mental from going the way of the miracle-performing angels, "stripped of their miraculous powers and governed by the same laws everything else is?"[209]

To the first question, I'm not sure how important the mere *duality* of the dualist's ontology ought to be. If, as the substance dualist believes, fundamentally mental substances—and properties, and events—are *real,* then what is to prevent them from eventually becoming the subject of a science? And what better science than the science of the relatively fundamental? If the physicalist is willing to accept a definition of "physical"

[208] Strictly speaking, P must also be *nonmiraculous*—or, compatible with scientific investigation. I discuss this shortly.

[209] (Dowell 2006, p.41)

that is compatible with the possibility of fundamental mentality, then the failure of dualism to entail the falsity of physicalism is hardly the fault of the dualist.

In "Why 'non-mental' won't work: on Hempel's dilemma and the characterization of the 'physical,'" Neal Judisch considers the possibility of mentality being incorporated by an idealized understanding of physics. He maintains that such a scenario would not, and ought not, trouble most dualists. On the contrary, he writes:

> It is in sensitivity to just this concern that most philosophers who deny that physics can account for a particular feature of the mental append to that judgment a proviso: ''absent a major revision in our conceptual repertoire'' or some ''revolutionary change'' in scientific theorizing, or what have you—after all, it would take an unusual degree of clairvoyance to know *a priori* that something like this couldn't take place, especially in the absence of any rough proposal about what the pertinent changes would have to be.[210]

Judisch's point is a simple one: we don't know what the physicists will discover. For that reason, a dualist who is committed to the existence of mental entities that are quite unlike the entities postulated by *present day* physics might, nevertheless, concede the possibility of entities of this sort being discovered by some *future, ideal* physics. Dualists are dualists because the monism presented by current physics strikes them as inadequate. I find it hard to believe that dualism *per se*—the commitment to there being *two kinds of things*—is of particular importance to most dualists.

What *is* of particular importance to most dualists is the preservation of certain features of the mental: intentionality, subjectivity, the presence of qualitative consciousness, and—for some, at least—freedom. Before simply accepting the possible physicality of the mental, then, a dualist ought first to be sure that the *second* objection can be met. Granting that Dowell's physicalism allows for the incorporation of mental

[210] (Judisch 2008, p312).

entities into the Physical_D, can these important *features* of mentality survive the incorporation? If not, then it is only nominally true that the mental is preserved.

I don't see why not. That is, I don't see anything in Dowell's account that would suffice to *prevent* these features from being preserved. To see why, consider again Dowell's response to the possibility of miracle-performing angels. She writes,

> A miracle-performing angel is an entity whose acts are by definition incapable in principle of being fit into a pattern of explanation characteristic of scientific theories. So if angels were to figure in our ideal physical theory…it would have to be in some mundane sense of 'angel'. They would have to be angels stripped of their miraculous powers and governed by the same laws everything else is. (41)

If, by "miracle," Dowell means "a violation of the laws of ideal physics," then clearly physicalism cannot allow for miracle-performing angels. Entities that violate the actual laws of a science cannot be captured by that science; the activity of such things could neither be predicted nor explained given the theoretical resources alone. If genuine mentality is like *this*, then, Dowell's physicalism cannot allow for the existence of the mental, properly speaking.

That said, I don't see why the dualist need insist that mentality is wholly anomalous, capable of law-breaking activity. True, the laws of *current* physics certainly can't capture all that there is to the mental, but why think those are all the laws that there are? Indeed, why assume that they are even *true*? If Dowell is to tie her physicalism to the future of the science of the relatively fundamental, then she has to accept the possibility of all manner of surprises. Just as there might be fundamentally mental *substances, properties, and events*, there might also be *laws* that govern those entities—laws with which we are presently unfamiliar. Perhaps those laws will be deterministic, mechanistic, and otherwise incompatible with much of what we take the mental to be—but perhaps

not.[211] They might, instead, be probabilistic laws, irreducibly intentional in nature and compatible with what we believe about the mental. That's the thing about the future, we don't know what will happen!

I think it would be helpful here to briefly consider a quotation from William Hasker's *The Emergent Self*. After detailing his own dualistic account, which includes the emergence of irreducibly mental substances from the physical world, Hasker considers the charge that these substance might just be physical after all. He responds as follows:

> If philosophers are prepared to stretch the meaning of 'physical' to encompass everything that has been said here about the field of consciousness, then so be it. What is *not* acceptable, however, is for someone to take the claim, thus arrived at, that 'the mind is physical' and use it as a premise from which to infer characteristics of the conscious mind that are contrary to the ones postulated in this [account].[212]

This is, I think, just exactly the right response. If the mental is real, and if it is fundamental, then any complete science that captures all that is fundamental will capture the mental—including whatever relevant laws there might be. By calling that science "physics," the physicist is *not* thereby entitled to infer that all that is fundamental is *just like the things we now call physical*.

Of course, all of this is highly speculative. Dowell would claim that all of this is highly *unlikely*; that, empirically speaking, there is just very little reason to believe that any of this will come about. That's alright, though. All that is needed here is possibility. Obviously, in order to justify her dualism, the *dualist* ought to believe it likely that fundamental mentality is real, and is quite unlike what we presently understand the physical to be. In order to respond to the *exclusion argument*, however, she need not

[211] (As evidence of this possibility, we might look to quantum physics, which has *already* surprised the physicists with the introduction of indeterminacy where we previously took determinism to be the rule.)

[212] William Hasker, *The Emergent Self*. (Ithaca and London: Cornell University Press, 1999), p.201

persuade the physicalist of the likelihood of the eventual discovery of fundamental mentality. For *this* purpose, the mere possibility suffices.

The causal closure principle under consideration, $closure_D$, is compatible with the possibility of a fundamentally mental cause being, at the same time, a physical cause. It is, therefore, *inadequate* as a premise in the causal exclusion argument, for it cannot support premise (3). Without premise (3)—without, that is, the claim that the mental cause is not identical to the physical cause required by *closure*—there is no exclusion to be had. $Closure_D$, like $closure_M$, fails as the crucial premise in the causal exclusion argument against substance dualism.

§6.4 The Causal Closure of the Physical$_W$

We are left, finally with Wilson's physicalism. Before attempting to formulate $closure_W$, I will simply state that, in contrast to the first two closure principles, $closure_W$ is neither obviously false nor compatible with the possibility of *sui generis* mental causation. (It is, however, a bit unwieldy as principles go!) By applying Wilson's definition of "physical" to *closure*, we get the following:

> $Closure_W$: If an *event that is a posit of the future (idealized) science of the relatively fundamental* **and** *is not fundamentally mental* has a sufficient cause that occurs at *t*, it has a sufficient cause that occurs at *t that is a posit of the future (idealized) science of the relatively fundamental* **and** *is not fundamentally mental*.

If $closure_W$ is true, then interactionist substance dualism is false. This closure principle explicitly precludes the possibility of fundamental mentality exerting causal influence in

the physical world. For that reason, it is certainly *strong* enough to serve as a premise in the causal exclusion argument.

However, for precisely that reason, it is altogether too strong to be of much use. If $closure_W$ is true, then interactionist substance dualism is false. *Clearly*, then, a substance dualist ought to reject it. After all, what is the justification for Wilson's closure principle? On what grounds are we to accept exclusion of the mental from the domain of the relatively fundamental? On *no* grounds, apart from the fact that fundamental mentality *just doesn't fit* with what physicalism has, historically, claimed exists. The extension of "physical" so as to include fundamental mentality would be *inappropriate*; the result would not, according to Wilson, be physicalism properly speaking. But notice that this is not a problem for the dualist. On the contrary, the dualist is *committed* to the claim that physicalism, understood so as to exclude fundamental mentality, is false. For that reason, no substance dualist should feel compelled to accept $closure_W$.

More significantly, no substance dualism *can* accept $closure_W$. It is incompatible with the possibility of interaction, and has, at its heart, the *stipulation* that dualism is false. If the causal exclusion argument is to be any threat to dualism, then, it cannot invoke, or otherwise rely upon, the truth of $closure_W$ Any argument against interactionist dualism that takes as a premise $closure_W$ begs the question, for $closure_W$ *just is* the denial of interactionist dualism.

§6.5 Final Thoughts on Closure

I have argued that none of the physicalist accounts that we have considered can support a causal closure principle that can be of use in a causal exclusion argument against interactionist dualism. On Melnyk's physicalism, *closure* is false. On Dowell's physicalism, *closure* is compatible with interaction. On Wilson's physicalism, *closure* begs the question that the causal exclusion argument purports to solve. None of these accounts, then, can provide what is needed in order to exclude mental causes from the physical world.

I hope now, briefly, to show that the problem lies not with these accounts in particular, but is rather a real, unavoidable result of Hempel's Dilemma. Consider the first horn of the dilemma—the claim that a physicalism based on *current* physics is almost certainly false. As far as I know, Melnyk is the only physicalist who has attempted to grasp this horn, and he does so only by *granting* the likely falsity of physicalism. As I hope to have shown, a physicalism that is (very probably) false yields a causal closure principle that is (almost certainly) false. No matter what the *physicalist* chooses to affirm, there is just no reason for a dualist to grant a principle that is so unlikely to be true. Unless the first horn can be grasped in such a way as to preserve the *truth* of physicalism, then, current-physics based physicalism can be of no use to the proponent of the causal exclusion argument.

There are, to be sure, more accounts that grasp the second horn than the two that I have here considered.[213] However, *all* of them must make the following choice: they can

[213] See, for example, J. Poland *Physicalism: the Philosophical Foundations*, (Oxford: Clarendon Press, 1994) Poland equates physics with the science of the occupants of space-time. If, as I have argued in Chapter Two, the occupants of space-time may very well include immaterial minds, this will also fail to

leave it to the physicists to tell us what will count as physical, and accept that the content of the concept can only be capture *a posteriori*, or they can choose, instead, to place *a priori* constraints on what can count as physical. If, like Dowell, they accept the former, they must accept the possibility of fundamental mentality. This possibility has not yet been ruled-out by the empirical science that is physics, and there is no good reason to assume that it *will* be ruled out. If, like Wilson, they accept the latter, then whatever *a priori* constraints they use to preclude the possibility of fundamental mentality will render their account precisely as question-begging as Wilson's is.[214]

Hempel's Dilemma is a significant challenge to physicalism, and it is an even greater challenge to anyone who would raise the causal exclusion argument against interactionist dualism. Neither the first horn nor the second allows for a causal closure principle that can be of any use in the argument. The dualist, then, really ought not to go to any lengths to render her dualism compatible with *closure*; depending upon what is meant by "physical," the principle is either *false*, *already compatible* with interaction, or *question-begging*.

---

exclude *sui generis* mental causes from the physical world. (In fn. 25, Dowell makes a similar, though not identical, point. She notes that Poland's account, unlike her own, would allow for miracle-performing angels, should they happen to occupy regions of space-time.)

[214] Question-begging, that is, *only insofar* as it is invoked in an *argument* against substance dualism. The physicalist is of course free to stipulate the falsity of interactionist dualism. What she may *not* do, of course, is to present those stipulations as *evidence against* interaction.

Conclusion

Despite its widespread acceptance, the claim that the physical world is causally closed is a dubious one. As we have learned from Hempel, the very notion of the physical world is far from clear, and without a consensus on what is meant by "physical" we can hardly claim consensus on what it means for the physical to be *closed*. Furthermore, as I hope to have shown, there is no good response to Hempel's Dilemma, at least not as it pertains to causal closure. On the first horn, *closure* is false; on the second horn, it is either too weak or too strong to be of any use. I know of *no* statement of closure that can suffice as justification for the premises of the causal exclusion argument against interactionist substance dualism.

For this reason, I suggest that the dualist *just stop worrying* about closure. This is not to say that she should reject causal closure, for as we have seen an outright rejection might not be necessary. Instead, the precise way that a dualist ought to respond to a causal closure principle will depend upon how that principle is disambiguated in light of the questions raised by Hempel's Dilemma. No matter how it is disambiguated, however, the result will not pose a real challenge to the dualist. For that reason, when responding to the causal exclusion argument, the dualist really would be better served by focusing her attention on *closure*, and not *exclusion*, as the problematic premise.

WORKS CITED

Alston, William "Perception and Representation" *Philosophy and Phenomenological Research* Vol. LXX, No. 2, March 2005 p.257

Bennett, Karen "Why The Exclusion Problem Seems Intractable, and How, Just Maybe, To Tract It," Nous 37:3 (2003) 471-497

Blackburn, Simon in "Supervenience Revisited." In Hacking, Ian (ed.) *Exercises in Analysis*. (Cambridge: Cambridge University Press, 1985) pp.47-67

Block, Ned and Robert Stalnaker "Conceptual Analysis, Dualism, and the Explanatory Gap." *Philosophical Review*, 108 (1), (1999): 1-46.

Boyd, R. (1983): On the Current Status of the Issue of Scientific Realism, *Erkenntnis* 17, 135 61-69.

------ (1985): 'Lex Orandi est Lex Credendi', in P.M. Churchland and C.A. Hooker (eds.), *Images of Science*, Chicago: University of Chicago Press.

Chalmers, David J *The Conscious Mind*: In Search of a Fundamental Theory. (New York & Oxford: Oxford University Press, 1996)

Dowell, J.L. "The Physical: Empirical, Not Metaphysical," *Philosophical Studies* (2006) 131:25-60, p.37

Dretske, Fred "Reasons and Causes," *Philosophical Perspectives* 3 : 1-15. 1989

Fodor, Jerry *A Theory of Content and Other Essays*. (Cambridge, Mass: Bradford Book/MIT Press, 1990)

Hasker, William *The Emergent Self*. (Ithaca and London: Cornell University Press, 1999), p.201.

Hempel, C. (1965): 'Aspects of Scientific Explanation', in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: the Free Press.

-------- (1966): *Philosophy of Science*, Edgewood Cliffs: Prentice-Hall

-------- (1969): 'Reduction: Ontological and Linguistic Facets', in Patrick Suppes, Sidney Morgenbesser and Morgan White (eds.), *Philosophy, Science, and Method: Essays In Honor of Ernest Nagel* (pp. 179–199), St. Martin's.

Hill, Christopher S. and Brian P. McLaughlin, "There Are Fewer Things in Reality Than Are Dreamt of in Chalmers's Philosophy." *Philosophy and Phenomenological Research,* 59(2), (1999): 445-454.

Horgan, Terry "Mental Causation and the Agent Exclusion Problem" *Erkenntnis: An International Journal of Analytic Philosophy*, 67(2), 183-200. September 2007

Jackson, Frank "Epiphenomenal Qualia," *Philosophical Quarterly* 32 (1982): 127-136.

Judisch, Neal "Why 'non-mental' won't work: on Hempel's dilemma and the characterization of the 'physical'" (Philososphical Studies (2008) 140:299–318)

Kenny, Anthony *Descartes* (New York: Random House, 1968)

Kim, Jaegwon "Events as Property Exemplifications," in M. Brand and D. Walton (eds.), *Action Theory*. (Dordrecht: Reidel, 1976) pp. 159-77

--------"Postscripts on Supervenience," *Philosophy and Phenomenological Research* 45 (1984): 153-176. Reprinted with permission in Kim, Jaegwon *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993)

--------"'Strong' and 'Global' Supervenience Revisited," in *Philosophy and Phenomenological Research* Vol. 48, No.2, (Dec 1987) p.315

-------- "Mechanism, Purpose, and Explanatory Exclusion," *Philosophical Perspectives* 3 (1989): 77-108.

-------- "Concepts of Supervenience," *Philosophy and Phenomenological Research* 45 (1984): 153-176. Reprinted with permission in Kim, Jaegwon *Supervenience and Mind*. (Cambridge: Cambridge University Press, 1993)

-------- *Philosophy of Mind* (Boulder: Westview Press, 1996, p.147)

-------- *Physicalism or Something Near Enough*. (Princeton: Princeton University Press, 2005)

-------- "Making Sense of Emergence," *Philosophical Studies*, 95 (1999): 3-36

-------- *Mind in the Physical World*. (Cambridge: MIT, 2000)

--------*Physicalism, or Something Near Enough* (Princeton: Princeton University Press, 2005)

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

Levine, Joseph "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly* 64 (1983): 354-361

------- *Purple Haze* (Oxford: Oxford University Press, 2001)

Lowe, E.J. "Non-Cartesian Substance Dualism and the Problem of Mental Causation," *Erkenntnis* (2006) 65:5–23

------- "Causal Closure Principles and Emergentism" *Philosophy*, 75, 571-585 (2000).

------- "Physical Causal Closure and the Invisibility of Mental Causation" in Sven Walter and Heinz-Dieter Heckmann, eds. *Physicalism and Mental Causation*. (Exeter: Imprint Academic, 2003)

McGinn, Colin "Another Look at Color," *Journal of Philosophy* Vol.93, No.11 (Nov. 1996) pp 537-553

Melnyk, Andrew "How to Keep the 'Physical' in 'Physicalism'" *The Journal of Philosophy*, Vol. 94, No. 12 (Dec., 1997), pp. 622-637; p. 622

------- "Some Evidence for Physicalism," in Sven Walter and Heinz-Dieter Heckmann, eds. *Physicalism and Mental Causation*. (Exeter: Imprint Academic, 2003)

-------*A Physicalist's Manifesto*. (Cambridge: Cambridge University Press, 2003)

Mills, Eugene "Interaction and Overdetermination," *American Philosophical Quarterly* 33.n1 (Jan 1996): pp105(13).

Montero, Barbara "Varieties of Causal Closure" in (Walter & Heckmann 2003), pp. 173-190.

Nagel, Ernest *The Structure of Science*. (New York: Harcourt, Brace and World, 1961)

Nagel, Thomas "What Is It Like To Be a Bat?" *Philosophical Review* 83 (1974): 435-450

Ney, Alyssa "Can an Appeal to Constitution Solve the Exclusion Problem" *Pacific Philosophical Quarterly*, 88(4), 486-506. December 2007

Papineau, David *Thinking About Consciousness*. (Oxford: Oxford University Press, 2002)

Plantinga, Alvin "Evolution, Epiphenomenalism, Reductionism" in *Philosophy and Phenomenological Research*, 68 (2004): 602-619

Poland, J. *Physicalism: the Philosophical Foundations*, (Oxford: Clarendon Press, 1994)

Putnam, Hilary "Psychological Predicates." in W.H. Capitan and D.D. Merrill (eds.), *Art, Mind, and Religion*. (Pittsburgh: University of Pittsburgh Press, 1967)

Sellars, Wilfred *Science, Perception and Reality*. (London: Routledge and Kegan Paul, 1963)

Van Inwagen, Peter *An Essay on Free Will*. (Oxford: Oxford University Press, 1983)

Warfield, Ted and Thomas Crisp "Kim's Master Argument" Nous 35:2 (2001) 304–316

Wilson, Jessica "On Characterizing the Physical." (*Philosophical Studies* (2006) 131:61–99)

Yablo, Stephen "Mental Causation" *Philosophical Review* 101, pp. 245–280. (1992)

Zimmerman, Dean "Material People" in Michael J. Loux, Dean W. Zimmerman, eds. *The Oxford Handbook of Metaphysics* (Oxford: Oxford University Press, 2003) p.492