

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]

UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MINDS, CAUSES AND THE EXCLUSION ARGUMENT

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

By

EDWARD COX

Norman, Oklahoma

2000

UMI Number: 9975798

UMI[®]

UMI Microform 9975798

Copyright 2000 by Bell & Howell Information and Learning Company.

**All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.**

**Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346**

© Copyright by EDWARD COX 2000

All Rights Reserved

MINDS, CAUSES AND THE EXCLUSION ARGUMENT

**A Dissertation APPROVED FOR THE
DEPARTMENT OF PHILOSOPHY**

BY

Reinaldo Elugardo
Reinaldo Elugardo
Dissertation Supervisor

Monte Cook
Monte Cook

James Hawthorne
James Hawthorne

Chris Swoyer
Chris Swoyer

Francis Durso
Francis Durso

**In Memory of
Marvin and Beatrice Lowe**

CONTENTS

Abstract	viii
Chapter	
1. INTRODUCTION	1
1.1. Introduction	
1.2. Outline of the Work	
1.3. Assumptions of the Work	
1.4. Brief Historical Introduction	
 PART I. THE EXCLUSION ARGUMENT 	
Overview	24
2 THE EXCLUSION ARGUMENT	26
2.1 Introduction	
2.2 The Exclusion Argument and Substance Dualism	
2.2.1 Causal Closure of the Physical	
2.2.2 Physical Causation of the Mental	
2.2.3 Causal Exclusion Principle	
2.3 The Exclusion Argument and Constitution Physicalism	
2.3.1 Causal Closure of the Physical	
2.3.2 Mereological Supervenience	
2.4 The Exclusion Argument and Nonreductive Token Physicalism	
2.4.1 Token Physicalism and Event Identity	
2.4.2 <i>EANTP</i> and <i>Quausation</i>	
2.4.3 Generative/Determinative Exclusion Principle	
2.4.4 Causal Closure of the Physical	
2.4.4 Property Supervenience	
2.5 Conclusion	
3 VARIANT PREMISES OF THE EXCLUSION ARGUMENT	58
3.1 Introduction	
3.2 Supervenience of the Mental on the Physical	
3.3 Commitments of Supervenience	
3.4 Entailments among Variant Premises	
3.4.1 Causal Powers	
3.4.2 Causal Closure of the Physical	
3.4.3 Thesis of Supervenience of Properties and Causal Closure of the Physical entail thesis of Supervenience of Causal Powers.	
3.4.4 Thesis of Supervenience of Causal Powers entails thesis of Causal Closure of the Physical	
3.4.5 Thesis of Supervenience of Causal Powers does not entail thesis of Property Supervenience	
3.5 Conclusion	

PART II: ARGUMENTS AGAINST EPIPHENOMENALISM	
Overview	86
4 BRIEF ARGUMENTS AGAINST EPIPHENOMENALISM	90
4.1 Introduction	
4.2 Argument from Ordinary Language or Common Practice	
4.3 The No-Causation-at-All Argument	
4.4 Fodor's End-of-the-World Argument	
4.5 Davidson's Argument from Explanation	
4.6 Argument from Simplicity	
4.7 Arguments against Coherence of Epiphenomenalism	
5 KNOWING NONCONSCIOUS MENTAL ITEMS	109
5.1 Introduction	
5.2 Existence of Nonconscious Mental Entities.	
5.3 Block's Theory	
5.4 Knowledge of the Mental as Second-Order Properties	
5.5 Conclusion	
6 FIRST-PERSON KNOWLEDGE OF CONSCIOUS STATES	125
6.1 Introduction	
6.2 Problems with (1): Having Qualia	
6.3 Problems with (2): First-Person Knowledge of Qualia	
6.3.1 Privileged Access and Skepticism	
6.3.2 Acquaintance with Qualia	
6.4 Problems with (3): Qualia as Representations	
6.5 Problems with Reference to Qualia	
6.6 Objection to the Nomic Correlation Response	
6.7 Type-E and Token-E Revisited	
6.8 Conclusion	
PART III. PREMISES OF THE EXCLUSION ARGUMENT	
Overview	152
7 SUPERVENIENCE AND MENTAL CAUSATION	155
7.1 Introduction	
7.2 Criticism of Supervenience	
7.3 Consciousness	
7.4 Quantum Holism	
7.5 Individuation by Context	
7.6 Some Inadequate Solutions	
7.6.1 Burge's Solution	
7.6.2 Baker's Solution	
7.6.3 Fodor's Solution	
7.7 Wide Content and Wide Behavior	
7.8 Conclusion	
8 THE EXCLUSION PRINCIPLE	196

8.1	Introduction	
8.2	Yablo on Mental Causation	
8.2.1	The Problem of Mental Causation	
8.2.2	Yablo's Theory	
8.2.3	Criticism of Yablo	
8.2.3.1	Content Properties Failure to Supervene	
8.2.3.2	Physical Properties without Corresponding Mental Properties	
8.4	A New Exclusion Principle	
8.5	Mental Causation	
8.6	Objections	
8.6.1	Novel Causal Powers	
8.6.2	What God Needs to Do	
8.6.3	Epiphenomenal Causation	
8.6.4	Second Class Causation	
8.6.5	Extrinsic Causal Relevance	
8.7	Conclusion	

Works Cited	234
-------------	-----

DISSERTATION ABSTRACT

The problem of mental causation, in the form of the Exclusion Argument, affects any variety of dualism or Nonreductive Physicalism. I argue for a solution to the problem of mental causation for Nonreductive Physicalism.

The Exclusion Argument consists of a broadly physicalist set of theses, with a different variant for each variety of dualism. These theses are the following:

- 1) *Dualism*. The mental and physical are real.
- 2) *Irreducibility*. The mental is not type-identical to the physical.
- 3) *Supervenience*. The mental is determined by the physical.
- 4) *Causal Closure of the Physical*. The physical is closed to causal influence from anything nonphysical.
- 5) *Exclusion Principle*. If there is one complete, independent cause (or causally relevant property), then there is no other cause (or causally relevant property).
- 6) Therefore, *Epiphenomenalism*. The mental is causally irrelevant.

The conclusion, Epiphenomenalism, follows from these broadly physicalist assumptions. The physical is determined only by the physical, so the mental cannot determine the physical. The mental is also completely determined by the physical, so the mental cannot determine the mental. Thus, the mental has no causal work to do. Since the physical completely determines everything, the mental can determine nothing. This sketch is the essence of any version of the Exclusion Argument.

I set up the most perspicuous version of the Exclusion Argument for the purpose of determining which premises are necessary for the argument and finding any important relations among the premises assumed in different versions of the argument. In this way I avoid focusing on questions irrelevant to the core issues of the Exclusion Argument.

Next I argue that the mental is causally relevant, so the Exclusion Argument is unsound. Epiphenomenalism is not readily shown to be false, as some philosophers would claim. However, Epiphenomenalism raises difficult problems of first-person knowledge and reference that suggest that this view is false. For this reason, either at least one of these principles of Nonreductive Physicalism must be false or the reasoning from these premises to the conclusion must be invalid.

Finally, I consider the most plausible attempts to solve the problem of mental causation. The most plausible theories of mental causation assume a dual-explanandum or two explananda view, but such views are inconsistent with the Exclusion Argument. To make such a theory possible, one must reject either the supervenience of the mental on the physical, the Exclusion Principle, or one must reject the reasoning of the Exclusion Argument.

Many philosophers doubt the premise that local physical properties determine the mental. The importance of context in the individuation of content properties, and perhaps special science and mental properties, requires rejection of the determination of the mental by local physical properties. However, the failure of supervenience can only allow for the mental to be causally relevant if some mechanism allows for mental causation. I argue that no such mechanism of mental causation is compatible with failure of supervenience.

The premise that leads to the problem for Nonreductive Physicalism is the Exclusion Principle itself. The Exclusion Principle, naively conceived, is false. The naive Exclusion Principle excludes as causally irrelevant events or properties that

intuitively appear to be relevant. For example, the Exclusion Principle incorrectly excludes determinables as causally irrelevant in favor of their determinates. A traffic light's being *scarlet* causes one to stop at the light, but this fact does not show that light's being *red* is not causally relevant to one's stopping. Some sets of events or properties do not compete for causal relevance. This fact contradicts the naïve Exclusion Principle. Thus, if the mental depends on the physical in the correct way, the mental would be incorrectly excluded from causal relevance by the naïve Exclusion Principle.

These counterexamples suggest a more sophisticated Exclusion Principle, but the more sophisticated principle does not rule out causal relevance for the mental if the mental is dependent upon the physical. Thus, I propose that the mental closely depends on the physical, and this dependence allows for mental causation consistent with a correct Exclusion Principle. Thus, the Exclusion Argument, including the sophisticated Exclusion Principle as a premise, is invalid, and mental causation is shown to be possible for Nonreductive Physicalism.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Understanding how the mind fits into a physical world is the philosophical mind-body problem. Some features of the mind, for example intentionality and consciousness, make it difficult to see how the mind could be part of a purely physical world. In contrast, the problem of mental causation, the problem of making room for the mental in the causal order of the world, makes it difficult to see how the mind could be anything besides a purely physical phenomenon. In particular, one variety of the problem, or perhaps one part of a larger problem, is presented by the Exclusion Argument. The Exclusion Argument starts from certain plausible views about the structure of the mental and the world to reach the conclusion that the mental cannot have causal power.

Descartes¹ famously encountered the problem of mental causation because of his view of the mind as an immaterial substance that somehow interacted with the body. Descartes believed that the mind was a non-physical substance whose essence was to think or be conscious whereas the body was a physical substance whose essence was to be extended in space. Critics of Descartes, even from his own time, have thought that this distinction between the mind and the body, if true, would make it impossible for them to interact. How, one might ask, could the non-physical mind 'push' on the physical body to cause it to act in the way one wanted it to? How could brute matter have any effect on a non-physical mind? Descartes was unable to explain the interaction adequately. Descartes'

¹ Whether Descartes' real problem is the one characterized here is a matter for scholars of Descartes. Given his belief that God constantly recreated the world, it is hard to see how his view allows for anything but God to cause anything. Descartes' real problem may thus have been one of how any substances could

problem of mental causation, then, was one of determining how non-physical mind-stuff might affect the body and, of historically less importance, how the body could affect the mind.

Physicalism easily solves Descartes' problem of mental causation. Indeed, a large part of the motivation for accepting Physicalism is its promise of solving the problem of mental causation. If the mind and body are both physical, the mind can affect the body just as any other physical object could. If both the mind and body are physical, the mind and body can 'push' on each other in any way required. Physicalism, thus, promises and is motivated by this promise of a solution to the problem of mental causation.

Nonreductive Physicalism, the current conventional wisdom in philosophy of mind, ironically, has its own version of the problem of mental causation. Nonreductive Physicalism in general is the view that there are no immaterial substances; all that exists is physical or an aggregate of the physical. There are two varieties of Nonreductive Physicalism. One variety of Nonreductive Physicalism is the view that the mental is token-identical but not type-identical to the physical. My instance of pain may be a brain state, but not all instances of pain are brain states since it is possible that creatures that feel pain not have brains like ours. As Putnam (1967) first pointed out, octopi certainly feel pain but have brains much different from ours. Martians or aliens, if such beings are possible, may have silicon brains that are completely unlike ours physically, but these beings may still be capable of feeling pain. So, the mental state-type *pain* is not the same as the brain state-type *C-fiber firing* or whatever it is that occurs in humans when they feel pain. One could

interact at all. Nonetheless, the problem as given above represents the traditionally received view of Descartes and presents a major obstacle for anyone advocating substance dualism.

somewhat contentiously describe this view is as materialist monism combined with property dualism.

The other variety of Nonreductive Physicalism is that the mind is not even token-identical to the physical but is only constituted by or composed of the physical. The mind, on this view, is a macroscopic object, say a brain, composed of smaller parts, and this object is not identical to the sum of its parts. Both nonreductive physicalist views entail that the mind is composed of nothing but physical substance, but mental types are not identical to any physical type.

The problem of mental causation now arises in a different guise. Mental events can unproblematically cause physical events if mental events are physical events. However, since each mental event has distinct mental and physical properties, it becomes legitimate to ask whether it was in virtue of the mental property of the cause that the effect occurred or whether it was in virtue of the physical property of the cause that the effect occurred. For example, consider my reaction to touching a hot stove. The brain event caused by my touching the stove causes me to pull my hand away. The hand and brain are connected by a series of neurons. If any form of Physicalism is true, these neurons from the hand to the brain and back constitute a complete causal chain from my touching the hot stove to a series of neural events in my arm and brain to the physical event of my hand pulling away. If, as Physicalism claims, that brain event is a mental event, the mental event causes me to pull my hand away. However, since the brain event's being a pain is not the same as its being, say, a C-fiber firing, one may legitimately ask whether my pulling my hand away occurred in virtue of the brain event's being a pain or its being a C-fiber firing. According to the neural story told above, it seems that my pulling my hand away occurred in virtue of the event's

being a C-fiber firing. And, since the neural causal chain is complete, my pulling my hand away did not occur in virtue of anything else. Assuming that anything unnecessary for this causal chain is irrelevant, my pulling my hand away occurs not in virtue of the cause's being a pain but in virtue of its being a C-fiber firing, and, generalizing, mental properties are causally irrelevant or epiphenomenal. If this line of reasoning is sound, the problem of mental causation reappears for Nonreductive Physicalism in the form of Epiphenomenalism of mental event *types*.

In some respects, Nonreductive Physicalism is more problematic than Cartesian Substance Dualism. Substance Dualism at least allowed mental events to cause other mental events, even if it seemed impossible for the mental and the physical to interact. However, Nonreductive Physicalism appears to have the consequence that mental properties cannot be relevant even for other mental properties.

The reasoning quickly sketched above for the case of pain is the essence of the Exclusion Argument. Nonreductive Physicalism's metaphysical commitments seem to entail that the mental is causally irrelevant. In the next chapter I will explicate varieties of this basic argument schema, but in schematic form the Exclusion Argument makes use of the following nonreductive physicalist premises².

Exclusion Argument Schema

1. *Dualism*. Mental and physical items are both real.
2. *Irreducibility*. No mental item is identical to any physical item.
3. *Supervenience*. Mental items supervene on or are determined by physical items.
4. *Causal Closure of the Physical*. No physical effect has a nonphysical cause.
5. *Exclusion Principle*. Any effect can have only one complete, independent cause.
6. Therefore, *Epiphenomenalism*. No mental item causes any effect.

² This argument uses the term 'item' to refer to any entity, event, state, process, or property; it uses the term 'cause' to refer to any cause or causally relevant event, state, process or property; and it uses the term 'effect' to refer to any effect or mental property that might be brought about by an event, state, process or property.

I will explain how the Exclusion Argument works roughly as follows. This explanation of the argument involves several oversimplifications that I will address in Chapters two and three of this work. A graphical representation of a putative case of mental causation should make the argument clearer.

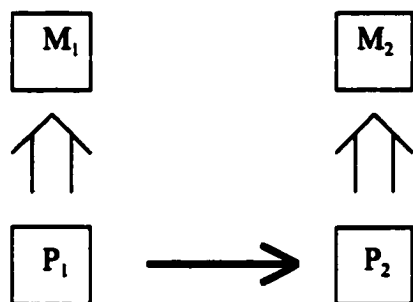


Figure 1

Figure 1 represents what occurs in a typical case of putative mental causation. M_1 and M_2 represent real mental items, perhaps including behavior. We might take M_1 to be an instance of pain and M_2 to be an instance of a propositional attitude token, my belief that I am in pain. P_1 and P_2 represent real physical items. Let's say for the sake of the example that P_1 is a brain event, following the standard oversimplified philosophical example a C-fiber firing, and P_2 is a further brain event, one that correlates with my belief-token that I am in pain. The single arrow represents causation. The double arrows represent determination or supervenience with the physical items as the determinants or supervenience bases and the mental items as the dependent or supervenient items.

Now consider that the physical item P_1 causes P_2 , and, by the Causal Closure of the Physical, principle 4, M_1 cannot cause P_2 . The C-fiber firing causes the brain firing correlated with my belief, and my pain cannot cause this brain event. Thus, the mental item M_1 , in this case, my pain, cannot cause P_2 . Figure 2 includes an arrow with a line through it representing this fact.

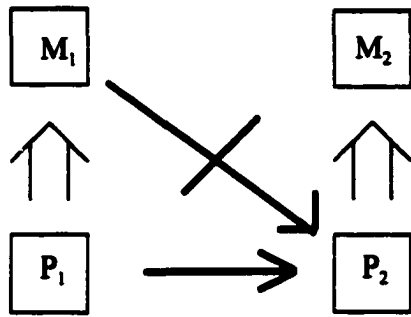


Figure 2

Moreover, M_1 cannot cause M_2 since P_2 already has a cause. According to the Supervenience thesis, premise (3) of the Exclusion Argument, P_2 determines M_2 . Since P_2 has P_1 as its cause, P_1 causes both P_2 and M_2 . The C-fiber firing causes the belief that I am in pain and the neural state that correlates with this belief. And according to the Exclusion Principle, premise (5) of the Exclusion Argument, when there is one cause of an event, there cannot be another. So M_1 cannot cause M_2 ; my pain cannot cause my belief that I am in pain. I represent this in figure 3 by drawing an arrow with a line through it between M_1 and M_2 .

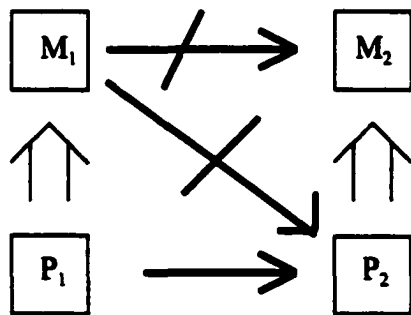


Figure 3

Since M_1 cannot cause either M_1 or P_2 , there is nothing that M_1 might cause. Therefore, Epiphenomenalism is true; mental events, like my pain M_1 , cause nothing.

One simplifying assumption I have made in presenting this argument is that there is only one immediate cause for any effect. In fact, there are an indefinitely large number of causal conditions for any effect, any one of which might be designated the cause depending

on our interests. A gas leak in one's home might be considered the cause of an explosion, and one's lighting a cigarette might be considered only a causal condition of the explosion. But, for someone working at the natural gas company, lighting a cigarette might be considered the cause and the presence of natural gas only a causal condition of the explosion. In each case the causal conditions, the presence of natural gas and the person lighting a cigarette (among other factors), are the same, but what we designate *the* cause differs according to our interest. By "cause," then, for the rest of this work, I will mean the complete set of causal conditions that bring about an effect.

The conclusion of the Exclusion Argument is Epiphenomenalism, the view that the mental does no causal work. In addition Epiphenomenalists see the mental as parallel to the physical and dependent on the physical. To take an example from Huxley (1898), epiphenomena are like the steam whistle of a train engine. The steam signals the working of the engine but has no effect on the machinery of the engine itself. Or to take the example of Plato's allegory of the cave, epiphenomena are like the shadows on the wall which appear to be causes but are only projections of the real causes occurring out of sight. These cases, the steam whistle and the projected shadows on the wall, provide examples of phenomena causally irrelevant to the mechanisms underlying them. These examples are not, of course, irrelevant to our perceptions. According to Epiphenomenalism, the mental is not only causally irrelevant to underlying mechanisms but is causally irrelevant *simpliciter*--it has no causal relevance whatsoever, not even to our perceptions.

The consequences of accepting Epiphenomenalism seem devastating. Causal relevance of the mental is necessary for us to be agents. If the mental is causally irrelevant, then we do not act but instead merely undergo processes. We would no more act when our

brains cause our movements than we act when our stomachs digest food or when our hearts pump blood.

Moreover, if the mental is causally irrelevant, then free will becomes impossible. For any theory of free will, it must be possible for us to act, even if, as Compatibilism claims, our actions are completely determined. Compatibilism requires as part of free action that we act, that our volitions, desires or some other mental state or event cause our actions. Libertarianism requires more of free action than merely that we cause our actions, but our causing our actions is an obviously necessary condition. So, the problem of mental causation negates our conception of ourselves as free and morally responsible agents. The incompatibility of Epiphenomenalism with agency is not a response to the Exclusion Argument, but it shows what is at stake and provides some compelling reason to find a flaw in the Exclusion Argument.

The problem of mental causation and the Exclusion Argument in particular is, in my view, the single greatest difficulty for Nonreductive Physicalism to overcome. Some philosophers³ have thought the argument sufficient to refute Nonreductive Physicalism, and thus to force a reconsideration of type Physicalism. The multiple realization argument, as noted above, resulted in a widespread rejection of Type Physicalism, but these new type physicalists⁴ believe that type Physicalism's demise was premature and that it should be accepted over Nonreductive Physicalism. The present work, then, is primarily a response on the part of Nonreductive Physicalism to the problem of mental causation presented by the Exclusion Argument but secondarily a defense of the nonreductive orthodoxy against type

³ See primarily Kim (1993b). Kim's recent work (Kim 1997 and 1998) favors a variety of analytical functionalism, but the skeptical challenge presented by the Exclusion Argument remains important in general and to his work in particular.

⁴ Including e.g. Bickle (1992) and some parts of Churchland (1984).

Physicalism. If Nonreductive Physicalism can allow the mental to be causally relevant, then this reason to reject the current orthodoxy in favor of a return to type Physicalism can be avoided. Thus, this work in part defends Nonreductive Physicalism by attempting to find a way for the mental to be causally relevant without accepting type Physicalism.

The remainder of this chapter will outline the goals of the later chapters of this work, mention some assumptions that I will make, and provide a brief historical introduction to the problem of mental causation and the Exclusion Argument. In the course of this historical introduction I will mention two other problems of mental causation that I will not pursue.

1.2 Outline of the Work

Chapters two and three of this work will constitute an extended look at the Exclusion Argument. Chapter two will discuss the premises of the Exclusion Argument in greater detail. Specifically, I will devise three different Exclusion Arguments for three different metaphysical positions. Chapter three will discuss some alternative premises often found in the literature and will give reason to reject them in favor of the schematic Exclusion Argument given here.

Chapters four, five and six will constitute a distinct section of this work that will argue at some length that we cannot accept Epiphenomenalism and so must instead reject the Exclusion Argument, either by rejecting one of the premises or by finding the Exclusion Argument to be invalid. Recently some philosophers (Bieri 1992, Chalmers 1996, and Horowitz 1999) have argued that Epiphenomenalism is more acceptable than is commonly believed. Chapter four of this work will consider and reject several arguments found in the literature on mental causation. Chapter five and chapter six will argue that, although

Epiphenomenalism is not easily refuted, nonetheless it does not accord with our knowledge of and ability to refer to our own minds.

Finally, chapters seven and eight will address the two most controversial premises of the Exclusion Argument. Chapter five argues that although at least some recent attacks on the view that the mental supervenes on the physical are likely correct, this failure of supervenience does not provide a sufficient explanation or mechanism for the mental to be causally relevant. The conclusion of chapter five is that the failure of supervenience alone cannot plausibly make mental causation possible.

Chapter six argues that the Exclusion Argument is either valid but unsound because the Exclusion Principle is false or invalid because the Exclusion Principle is too broad to exclude mental causation. The relation of the mental to the physical must be close for the mental not to be excluded. Thus, the final conclusion of this work is that the relation of the mental to physical must be a close dependence relation such that the Exclusion Principle does not apply to the case of the mental and the physical.

1.3 Assumptions of the Work

Before presenting the Exclusion Argument at length in chapters two and three, I will clarify some assumptions of the argument. First, I will assume that properties can be causally relevant. Second, I will assume that there is more to causation and causal relevance than regularities or counterfactuals. Third, I will assume some variety of scientific or explanatory realism.

I will assume that properties are causally relevant. There are three different varieties of causal relevance. The first variety is for an event to be causally relevant by causing an effect. Imagine, for example, the event of Hurricane Mitch striking Honduras. If Hurricane

Mitch is a purely physical event, or set of physical events, then Hurricane Mitch's striking Honduras can cause the damage to Honduras. Hurricane Mitch can be causally relevant simply by being a cause. The first, and least controversial, case of causal relevance is of an event causing an effect.

The second variety of causal relevance is for a property to be that property *in virtue of which* or *because of which* an effect occurs. Consider again the example of Hurricane Mitch striking Honduras. Let us say that Mitch is reported on page 13 of the New York Times. Mitch did a tremendous amount of damage to Honduras, but it was not in virtue of Mitch's being reported on page 13 of the Times that it damaged Honduras. Instead, Mitch damaging Honduras occurred in virtue of the high winds that made up the hurricane. Intuitively, some properties of Mitch were causally relevant to Mitch's damaging Honduras, but others were not⁵. Thus, properties can be responsible for a cause having the effect that it does although not all properties are relevant to all effects.

The third variety of causal relevance is for an effect to have a property in virtue of its cause having a particular property. This variety of causal relevance or *qua* causation as Horgan (1989) characterizes it, "c *qua* F causes e *qua* G." (Horgan, 1989, 50) To illustrate this notion Horgan provides a *qua* causation or *qua* locution paraphrase of statement (7),

(7) He exercised because he wanted to reduce and thought exercise would do it.
(Horgan, 1989, 49)

The paraphrase becomes statement (8)

(8) His exercising, *qua* the property *being an exercising* was caused by his desire to reduce and his belief that exercising would result in reducing, *qua* the respective properties *being a desire to reduce* and *being a belief that exercising would result in reducing*. (Horgan, 1989, 50)

⁵ The point cannot be avoided by interpreting the example as a mereological sum of microscopic events. The point would hold just as well if the example were one of an electron and its charge. I use the Hurricane Mitch example because of its intuitive appeal.

Causation is necessary for quausion on this view, but quausion requires more. Quausion requires that the effect, insofar as the effect is of the type that it is, occurs in virtue of the cause, insofar as the cause is of the type it is.

The essential point here does not require a property ontology. The point could be equally well put in terms of event types. Following McLaughlin (1989) one might characterize Epiphenomenalism (Type Epiphenomenalism) as the view that events occur in virtue of falling under physical types but not in virtue of falling under mental types. I will assume at least that events occur in virtue of falling under a type, or in virtue of being something of that type.

I assume that it is completely uncontroversial that causes are causally relevant. These second two types of causal relevance are more controversial. I can provide some reason to think that both of them are real features of the world, but a detailed defense of such a position would take us too far afield.

One argument for these latter two notions of causal relevance is to observe that merely giving the cause of the damage to Honduras is an incomplete explanation. A complete causal explanation would advert to a law that governs events of one type when they interact with events of another type. Thus, a complete explanation of the damage to Honduras would appeal to laws of force and momentum and to the properties of the hurricane that are adverted to in those laws. Less tendentiously we can always ask about the event of Hurricane Mitch striking Honduras why that event should be responsible for the damage to Honduras. Some property or set of properties of the hurricane require that the hurricane will have that effect when the hurricane strikes Honduras. If there is no property or type in virtue of which the effect occurs, it seems that all causal relations are either

accidental or involve an appeal to brute facts. But the fact that some events cause other events cannot be merely a brute fact; these occurrences cry out for further explanation. These considerations, I hope, establish at least the *prima facie* plausibility of the claim that properties can be causally relevant.

I cannot here argue in more detail for this assumption, so any readers who disagree on this point need read no further. Moreover, the problem of mental causation would not affect Token-Identity Physicalism if properties are not causally relevant. However, I find the claim that properties are causally relevant so overwhelmingly plausible that I cannot accept this solution.

My second assumption is that there is more to causation than regularities or counterfactuals. This point is related to the previous one in that one might explain why one event causes another by appealing to regularities. Appeal to regularities or counterfactuals as analyses of causation, I assume, are inadequate. Of course, regularity or counterfactual theories can easily solve the problem of mental causation. Lepore and Loewer (1989), for example, solve the problem of mental causation by accepting a counterfactual notion of causation. If causation is just a regularity or counterfactual relation, mental properties can easily fit into these regularities as well as physical properties can, especially if mental properties supervene on physical properties. However, I take it that accepting regularity or counterfactual analyses of causal relevance is really to accept a substitute for real causal relevance. Philosophers have argued by counterexample against these theories of causation. Correlative effects of a common cause, for example, show that regularity theories or simple counterfactual theories cannot properly analyze our notion of causal relevance. Any simplistic analysis in these terms would also fail to distinguish epiphenomenal byproducts

from actual causes. Although some sophisticated analysis may avoid these obvious counterexamples, it seems to me that causation and causal relevance are part of the basic furniture of the universe and cannot be reduced to anything simpler. Since a detailed argument for this claim is beyond the scope of this work, readers who disagree need read no further and, furthermore, if they are nonreductive physicalists, need have no worries about mental causation.

My third assumption is some variety of scientific or explanatory realism. Instrumentalism, the view that explanations in science are merely convenient fictions, in contrast to scientific or explanatory realism, can accept that some explanation, either mental or physical, best fits the data and so is a superior explanation. However, Instrumentalism would have no problem with accepting one or many conflicting explanations that were useful for certain purposes. Instrumentalism would then have no problem of mental causation.

Instrumentalism comes in several varieties, so it is impossible to give an unequivocal instrumentalist solution to the problem of mental causation. One possible Instrumentalist view is to deny the reality of both the mental and the physical. Since neither explanation would be considered true, no reason would exist to see them as conflicting explanations. Since I will explicitly assume realism (or what I shall call 'dualism') for the mental and the physical for each of three different Exclusion Arguments, I will not discuss this possibility at any greater length here.

A second possibility is that the Instrumentalist would see mental explanations as objectively true since the mental is directly observed. No philosopher that I know of is an Instrumentalist about everything. Thus, a sense-data theorist or Berkeleyan Idealist, might

accept the reality of directly observed mental objects, but reject the reality of physical objects, objects external to the mind. This variety of Instrumentalist might believe that because our minds and their contents are directly observed by us, but the physical is observed only through our conscious states, only the mental is real. On this view mental explanation is objectively true, whereas physical explanation can only be a useful instrument but not strictly true or false. On such a view only mental causation is real; physical causation is only a convenient fiction that gives useful predictions. So, on this view, the mental cause takes precedence over the physical cause whenever there is a potential conflict. On either view of the reality of the mental and physical, Instrumentalism has no problem of mental causation.

My views on these issues are orthodox enough to require no detailed defense here. Although rejecting these assumptions would provide an easy solution to the problem of mental causation, rejecting any of them would, I believe, commit one to problems even more serious. In any case, rejection of causal relevance of properties, acceptance of regularity or counterfactual theories of causation, and acceptance of Instrumentalism would appeal to only relatively small parts of the philosophical community. So, the problem of mental causation, as it applies to the nonreductive physicalist, is quite a real concern for most philosophers. Any attempted refutation of the above views detailed enough to convince any philosopher attached to them would be outside the scope of this work, so I must settle for setting them to one side at the beginning.

1.4 Brief historical introduction.

The Exclusion Argument is a recent philosophical development, but it has deep roots in the historical problem of mental causation. Contrary to the modern Exclusion Argument,

philosophers have long thought that explanations of behavior require appeals to more than the merely physical but also to the mental. Perhaps the first appearance of this claim is Plato's *Phaedo* (1989). Plato, through Socrates, argues for the necessity of teleological explanations and for the inadequacy of Anaxagoras' physicalist explanations. Anaxagoras, Socrates states, claims that Mind is the cause of all things, but, in reality, Anaxagoras makes no appeal to mind at all. Socrates says,

As I read on I discovered that the fellow made no use of mind and assigned to it no causality for the order of the world, but adduced causes like air and aether and water and many other absurdities. It seemed to me that he was just about as inconsistent as if someone were to say, The cause of everything that Socrates does is mind--and then, in trying to account for my several actions, said first that the reason why I am lying here now is that my body is composed of bones and sinews, and that the bones are rigid and separated at the joints, but the sinews are capable of contraction and relaxation, and form an envelope for the bones with the help of the flesh and skin, the latter holding all together, and since the bones move freely in their joints the sinews by relaxing and contracting enable me somehow to bend my limbs, and that is the cause of my sitting here in a bent position. Or again, if he tried to account in the same way for my conversing with you, adducing causes such as sound and air and hearing and a thousand others, and never troubled to mention the real reasons, which are that since Athens has thought it better to condemn me, therefore I for my part have thought it better to sit here, and more right to stay and submit to whatever penalty she orders. Because, by dog, I fancy that these sinews and bones would have been in the neighborhood of Megara or Boeotia long ago--impelled by a conviction of what is best!--if I did not think that it was more right and honorable to submit to whatever penalty my country orders rather than take to my heels and run away. (Plato 1989, 80)

Although the context here is one in which Socrates attacks Anaxagoras as inconsistent, Plato clearly asserts that explanations of human behavior require appeals to more than the facts about one's physical constitution but also to one's desires and beliefs about what is best. Plato's approval of Anaxagoras' claims for the causal power of Mind shows that Plato is appealing to something he takes to be obvious, specifically, that our intentions cause our actions and that any purely physical explanation of our actions would be

inadequate. Although the physical explanations appealed to in the passage are comically inadequate and could be no more than a caricature of any modern physicalist's position, Plato's claim can be seen as a challenge to physicalist explanation and its correlative assumptions that form the basis of the Exclusion Argument. The issue we can see Plato as raising for philosophy is whether Physicalism can ever completely explain human behavior.

Although Plato raised the topic of the adequacy of physical explanations of behavior, it was not until Descartes that philosophers seriously considered the possibility of completely mechanistic explanations of behavior. Descartes believed that animal, but not human, behavior could be explained completely mechanistically. In fact, of course, Descartes thought humans were immaterial thinking substances which interacted with the physical body.

Because of this distinction between the mental and physical substances, the problem of mental causation originates with Descartes. In the *Sixth Meditation*, Descartes argues that the mind and body are distinct substances, distinct independent entities. He claims the essence of body is to be extended, and the essence of mind is to think.

Descartes gives some indication of the relation between the mind and body in his *Sixth Meditation*. He writes,

Nature also teaches me, by these sensations of pain, hunger, thirst and so on, that I am not merely present in my body as a sailor is present in a ship, but that I am very closely joined and, as it were, intermingled with it, so that I and the body form a unit. If this were not so, I, who am nothing but a thinking thing, would not feel pain when the body was hurt, but would perceive the damage purely by the intellect, just as the sailor perceives by sight if anything in his ship is broken. Similarly, when the body needed food or drink, I should have an explicit understanding of the fact, instead of having confused sensations of hunger and thirst. For these sensations of hunger, thirst, pain and so on are nothing but confused modes of thinking which arise from the union and, as it were, intermingling of the mind with the body. (Descartes 1984, 56)

Later in the Meditation he describes how the body affects the mind, however, in a way that seems to contradict this notion of the mind and body existing as a union. He writes,

[W]hen I feel a pain in my foot, physiology tells me that this happens by means of nerves distributed throughout the foot, and that these nerves are like cords which go from the foot right up to the brain. When the nerves are pulled in the foot, they in turn pull on inner parts of the brain to which they are attached, and produce a certain motion in them; and nature has laid it down that this motion should produce in the mind a sensation of pain, as occurring in the foot. (Descartes 1984, 60)

These two passages seem inconsistent. How could the mind be a unit with the body and yet have the mind only be connected to the body in the brain? How could this interaction take place? Others soon pressed him to explain this interaction.

Pierre Gassendi, for example, in his *Objections to Descartes' Meditations* raises these types of objections to Descartes' notion of an immaterial, unextended substance and its interaction with the body.

How can there be effort directed against anything, or motion set up in it, unless there is mutual contact between what moves and what is moved? And how can there be contact without a body when, as is transparently clear by the natural light, 'naught apart from body, can touch or yet be touched.' (Gassendi 1984, 237)

To this criticism, Descartes' response is only that, "it is not necessary for the mind itself to be a body, although it has the power of moving the body." (Descartes 1984, 266) This response is clearly inadequate. So Descartes has given this theory about the distinction of the mind and the body such that it is impossible for them to interact. His only response when questioned appears to be that they do interact.

Descartes is less dismissive of Princess Elizabeth⁶ when she voices similar concerns, but does not appear to answer her questions. Eventually Elizabeth's urging led Descartes to work on the *Passions of the Soul*, published posthumously. Descartes, in this work, reasserts his notion of the mind as existing in union with the body, but then, famously, argues that the

mind and body interact in the pineal gland. A fair criticism of this response is that it tells us where the interaction occurs but not how it occurs.

At some points Descartes appears to believe that the mind can affect the physical as long as the physical is small enough. Descartes believes that the nerves work by means of very small animal spirits. In the *Treatise on Man*, Descartes notes that the animal spirits, described as a “very fine wind, or rather a very lively and pure flame” (Descartes 1984, v.1, 100) and as “the finer parts of the blood” (Descartes 1984, v.1, 100), go into the tiny pores in the pineal gland and then are sent ultimately to direct the limbs. This description suggests that the soul, situated within the pineal gland, directly affects the movements of the animal spirits. And, somehow, these animal spirits are small enough for the immaterial mind to affect. It is as if Descartes thought that the mind could pull on physical levers provided they were small enough. This strategy, although absurd, has recently been pursued by a leading neuroscientist and by a leading physicist⁷. Perhaps this view of Descartes is a caricature based on some incomplete remarks in the *Treatise on Man*. In the *Passions of the Soul*, Descartes’ explanation is clearer but ultimately no more helpful.

In the *Passions of the Soul*, Descartes claims that the animal spirits are moved by the pineal gland, and the pineal gland itself is moved by the soul. He writes,

Let us therefore take it that the soul has its principal seat in the small gland located in the middle of the brain. From there it radiates through the rest of the body by means of the animal spirits, the nerves and even the blood, which can take on the impressions of the spirits and carry them through the arteries to all the limbs. (Descartes 1984, v.1, 341)

⁶ Correspondence with Elizabeth in Descartes (1984) volume I.

⁷ The neuroscientist Eccles (1994) argues that the mind affects the body altering the quantum field of certain parts of the brain called presynaptic paracrystalline vesicular grids. Eccles’s theory is that if he can find some part of the brain small enough for quantum effects occur, then he can avoid problems of conservation of energy and momentum. The mind, he thinks, can affect apparently random occurrences within these structures to give rise to macroscopic changes in the brain. The moral of this theory is that emphasis on the conservation laws is misguided. The physicist Penrose (1989) appears to follow a similar strategy.

Thus, the animal spirits act through the blood and nerves on the muscles to cause our actions.

These animal spirits are affected in some manner by the soul in the pineal gland or acting through the pineal gland. Later he writes,

[T]he mechanism of our body is so constructed that simply by this gland's being moved in any way by the soul or by any other cause, it drives the surrounding spirits towards the pores of the brain, which direct them through the nerves to the muscles; and in this way the gland makes the spirits move the limbs. (Descartes 1984, v.1, 341)

Apparently, on Descartes' more considered view, the pineal gland is moved by the soul, and the animal spirits are moved by the pineal gland. But this view of the interaction of the mind and body says nothing about how the immaterial mind affects the physical pineal gland. Thus, whichever interpretation of Descartes is correct, Descartes says at best where the interaction occurs but cannot explain how this interaction is possible.

Descartes' problem of mental causation has been well documented. However, the Exclusion Argument itself does not appear until some time later. Although scientists and philosophers must have routinely applied some variety exclusion principle to sets of data, the Exclusion Argument as it is applied to the mind or consciousness makes its first unofficial appearance in Huxley's (1898) argument for Epiphenomenalism. Certainly scientists have always thought that once they have given a complete explanation of an effect, they did not also need to perform more experiments to show that some other explanation was not also correct. For example, if one can explain lightning completely in terms of electrical discharges in the atmosphere, one does not need to study further to determine whether the gods were also involved. However, Huxley was the first to apply this principle to the human mind.

Huxley, following Descartes, argues that the actions of animals can be explained entirely in terms of the actions of their nervous systems but believes further that the actions of humans can be explained entirely in terms of their nervous systems. On his view, since there is a complete neural causal explanation for human behavior, the conscious mind cannot be causally responsible for any behavior. Huxley cannot deny the existence of consciousness—we are too familiar with it in our own case—but we can show that consciousness cannot be a cause of any behavior.

Huxley's argument differs from the current Exclusion Argument in that it does not generalize to any sets of properties or entities. Huxley discovered, he thought, that the neural explanation of behavior was sufficient, so he rejected any connection between consciousness and behavior. Nonreductive Physicalism, on the other hand, supposes that the problem posed by the Exclusion Argument is one for any pairs of event types or sets of properties, not only one of the relation between consciousness and the neurons that underlie it.

Norman Malcolm (1968) introduced the Exclusion Argument into recent philosophy. Malcolm argued that physical, or neural, explanations of behavior could not be complete. The neural explanation, if it were complete, would have to compete with the mental, or intentional, explanation of behavior. Since the intentional explanation of behavior was superior, the neural explanation could not be complete. Malcolm's view provides an interesting example of how philosophical fashions change. Currently the worry is that intentional explanation might be supplanted by neurological explanations, but at that time the worry was just the reverse. The essence of Malcolm's Exclusion Argument remains the

same as ours; one complete explanation or set of causes for an event excludes any other explanation⁸.

The Exclusion Argument because of its premises involving the adequacy of physical explanations for human behavior and the metaphysical primacy of the physical is the modern analogue of traditional problems of mental causation. Two other problems of mental causation have appeared recently in philosophy of mind. These problems are the problem of Causal Relevance of Content and the problem of Anomalism of the Mental. Davidson (1970) raises the latter problem. This problem hinges on assumptions about the normative and holistic character of human rationality and the actions it gives rise to, and assumptions about the nomic character of causality. This problem is quite serious, but it does not arise immediately from basic physicalist commitments as the Exclusion Argument does, and, arguably, relies on notions idiosyncratic to Davidson. Thus, the Exclusion Argument presents a more fundamental problem for Physicalism, and so I will not address Davidson's argument.

The problem of the Causal Relevance of Content will overlap my discussion in the last two chapters of this work. This problem arises from the assumption that causation is entirely local and recent arguments that mental content is relational or extrinsic to the mind. This argument, since it does not arise immediately from basic physicalist assumptions, is also less fundamental to Physicalism than the problem presented by the Exclusion Argument. Whatever the merits of these other problems for mental causation, it is important to keep them separate from the argument I address in this work, the Exclusion Argument.

⁸ The fact that Malcolm frames his argument in terms of counterfactuals mars his argument. I have already noted that I reject counterfactual accounts of causation.

The Exclusion Argument presents a problem that is more fundamental to Physicalism and more in line with historical problems of mental causation and mental explanation.

PART I:
THE EXCLUSION ARGUMENT

OVERVIEW

The problem of mental causation is that of determining how the mind can be causally relevant in a physical world. This problem has been derived from a variety of different arguments. One such argument, currently of much interest, is the Exclusion Argument. An Exclusion Argument can be constructed for any of several metaphysical positions. Moreover several different sets of premises for the Exclusion Argument have been adduced to show that the mental must be causally inert. I will analyze the Exclusion Argument to show how it affects different varieties of dualism, to show whether the argument is valid and whether any of its premises might be false. Because a solution to the problem might depend on accidental features of a particular version of the argument, the first part of this work will set up the Exclusion Argument by constructing the most perspicuous version of it.

The most perspicuous version of the Exclusion Argument will include all the premises necessary for the argument to be valid, will exclude any unnecessary premises and will make explicit any different assumptions that might be combined in other versions of the argument. In chapter two I will discuss three Exclusion Arguments, one for each dualist position. Each Exclusion Argument will involve variations on the same premises. In chapter three I will discuss assumptions made in some other philosophers' versions of the Exclusion Argument and relations among these different premises. This discussion will show which premises are required for the argument and how premises from different versions of the argument are related. The goal of subsequent chapters will be to analyze the

more controversial premises of the Exclusion Argument to see which premise or premises might be rejected and to investigate the possibility that the mental is causally inert.

CHAPTER 2

THE EXCLUSION ARGUMENT

2.1 Introduction

An Exclusion Argument can be raised for any variety of dualism. In this chapter I will describe three different types of dualism and give Exclusion Arguments for each of these dualisms. The purpose of this chapter is to get clear on what precisely is required for an Exclusion Argument to entail Epiphenomenalism and to show what metaphysical views are affected by an Exclusion Argument.

The Exclusion Argument, as I showed in the first chapter, appears to entail that the mental is epiphenomenal, but there are different varieties of Epiphenomenalism. If the mental is epiphenomenal, it is causally inert and plays no role in world or in other mental activities. The mental rides along with the physical as the froth rides along with the waves. In Huxley's (1898) simile, it is like the whistle of the steam engine that signals that the engine is working but plays no role in the movement of the engine. According to Epiphenomenalism, the mind is nothing more than a causally irrelevant byproduct of the brain. All the causal work is done by the brain while the mind's activities merely accompany those of the brain.

Philosophers often consider Epiphenomenalism to be a particular problem for meaning or content. Dretske (1989) uses the example of an opera singer who shatters a glass by singing. Her singing caused the glass to shatter in virtue of its frequency, say, a high C. Supposing that she sang in words, her words had meaning, but the meaning of her words could not cause the glass to shatter. The meaning is causally irrelevant to the glass shattering; the glass shattered in virtue of the frequency of her singing but not in virtue of

the meaning of her words. Thus, if a physical mechanism accompanies every event, meaning may be epiphenomenal.

C. D. Broad (1925) first noted a distinction between two varieties of causal inertness. These two varieties of causal inertness are called Token-Epiphenomenalism and Type-Epiphenomenalism.

McLaughlin (1989) characterizes the distinction as follows:

Type Epiphenomenalism (Type-E). (a) Events can be causes in virtue of falling under physical types, but (b) events cannot be causes in virtue of falling under mental types.

Token Epiphenomenalism (Token-E). (i) Physical events can cause mental events, but (ii) mental events cannot cause anything. (1989, 109-10)

The characterization of Token-E is self-explanatory. Mental events, either conceived as belonging to a distinct substance or as aggregates of physical entities but not identical to the sum of those entities, cannot be causes. Type-E, on the other hand, requires some explanation. Type-E is the view that the mental *qua* mental is not causally relevant, that an event insofar as it is a mental event is not relevant to the production of an effect. Only the physical properties of an event are causally relevant to the occurrence of an effect; the mental property are not relevant.

If mental events are a species of physical events, then they can unproblematically be causes. If a mental event is a physical event, and the physical event is a cause, then the mental event is a cause. However, a mental event can be a cause without the effect occurring in virtue of the event's being a mental event, without the mental property being causally relevant. Consider the following example¹. Let us say Hurricane Mitch, the hurricane that ravaged Central America, was reported on page 13 of the *New York Times*.

¹ This example is essentially an updated version of one given by Antony (1989) following Davidson (1980).

Mitch caused a tremendous amount of damage in Central America. However, the damage did not occur in virtue of Mitch's being reported on page 13 of the *Times* but in virtue of Mitch's being a powerful hurricane. That Hurricane Mitch was reported on page 13 of the *Times* is causally irrelevant to Mitch's effects on Central America. In the case of Epiphenomenalism, the physical properties of the cause are already causally sufficient for the effect, so mental properties appear quite unnecessary, and, thus, effects do not occur in virtue of the mental properties of any cause. Mental properties appear to be, like Mitch's being reported on page 13 of the *Times*, unnecessary or epiphenomenal byproducts of actually causally relevant properties. So, mental events can still be causes, just as Mitch is a cause, but their effects may occur only in virtue of their physical properties not in virtue of their mental properties. This view of mental causation is Type Epiphenomenalism or Type-E.

As we have seen, Type-E does not entail Token-E, but Token-E does entail Type-E. Type-E does not entail Token-E because mental event tokens may be identical to physical event tokens. In that case the mental would be Type epiphenomenal without being Token epiphenomenal; mental events would be causes, but nothing would occur in virtue of falling under a mental type. However, Token-E does entail Type-E. If mental events cannot cause anything, then no mental event can be a cause in virtue of its mental properties. If the mental event is not a cause, then no property of the event is causally relevant to producing any effect. So, Type-E could be true without Token-E being true, but Token-E could not be true without Type-E being true.

I take the *in-virtue-of* relation in this case to be a real relation of causal relevance. Because of its close similarity to the *because-of* relation, McLaughlin (1989) takes the

relation to be an explanatory and, thus, epistemic relation. On my view, the similarity of the in-virtue-of relation to the because-of relation is explained by the fact that the in-virtue-of relation is an objective relation that one might appeal to in explanations but which is independent of our explanations and other epistemic endeavors. In the case of Type-E, the in-virtue-of relation is one of causal relevance to an effect.

Even if one agrees with McLaughlin about the nature of this relation, assuming explanatory realism, as I noted in the previous chapter that I do, entails that some real relation grounds this epistemic one. Thus, I take the in-virtue-of relation to be a real relation or at least correspond to some real relation. Thus, although I disagree with McLaughlin about the nature of the relation, if we are to accept explanatory realism, then we must accept that some real, and not purely epistemic, relation is involved in Type-E.

Most contemporary philosophers of mind have thought it obvious that the mental has causal power. Virtually all people agree that their thirsts cause them to drink and their pains cause them to flinch. As Jerry Fodor writes in his inimitable style,

Whereas, 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. (1989, 77)

However, belief in the causal efficacy of mental properties is inconsistent with the fundamental notions to which the Exclusion Argument appeals. If we are to hold to the causal efficacy of the mental, we must find the argument to be unsound. Unfortunately, the argument appears to be valid, and each of the metaphysical assumptions is independently plausible. The Exclusion Argument, then, leads to the problem of mental causation. We think mental causation occurs, yet it cannot occur if this plausible metaphysical picture is

correct. The purpose of this work, then, is only in part to argue that the mental must have causal power but primarily to show how the mental can have causal power.

A variation of the Exclusion Argument afflicts every type of dualism. The Exclusion Argument is roughly the following. The mental and the physical are distinct and real existents. The physical determines the mental and the mental can have no effect on the physical. Because the mental is determined by the physical, each mental effect is completely determined by the physical. Finally, because there can be only one complete, independent determiner for any effect, the mental cannot be causally relevant to any mental or physical effect.

Three forms of the Exclusion Argument correspond to each of three forms of dualism. I will show how each of these arguments appears to show that the mental is causally inert on the given metaphysical view. I will also show, for each view, why we might think the premises for that argument are true. I will put off careful consideration of the contentious premises and validity of forms of the Exclusion Argument until later.

The three forms of dualism are Substance Dualism, Constitution Physicalism, and Token Physicalism combined with Property Dualism. Substance Dualism is the Cartesian view that there are two distinct types of entities or independent existents. Constitution Physicalism is a Physicalism in that it does not postulate immaterial substances but is a dualism of macro-entities and their microconstituents. Constitution Physicalism claims macroscopic events, states and objects are not identical to the mereological sums of their microphysical constituents. Token Physicalism is the view that every particular is physical but that event, state or object types are not the same as mental types. Each dualism is susceptible to a version of the Exclusion Argument

Schema given in chapter one.

2.2 The Exclusion Argument and Substance Dualism

Exclusion Argument for Substance Dualism (*EASD*)

- (1) *Substance Dualism*. The mental and the physical are real substances.
- (2) *Irreducibility*. Mental substances are not identical to physical substances.
- (3) *Physical Causation of the Mental*. The physical causally determines the mental. All mental events are caused by physical events.
- (4) *Causal Closure of the Physical*. There can be no cause of a physical event that is not itself a physical event.
- (5) *Causal Exclusion Principle*. There can be no more than one complete, independent cause for any event. If there is one complete, independent cause for an event, there cannot be another cause for that event, except in cases of overdetermination.
- (6) Mental causation is not a case of causal overdetermination.
- (7) Therefore, *Token-Epiphenomenalism*: the mental substance cannot be a cause of any effect.

The Exclusion Argument here follows the general pattern given in chapter one as the Exclusion Argument Schema. The assumptions of Substance Dualism and Irreducibility, premises (1) and (2), the mental and physical are distinct substances. Because of the Causal Closure of the Physical, premise (4), the mental substance cannot affect the physical substance. Thus, the mental is a distinct substance that cannot affect the physical. Because of the Physical Causation of the Mental, premise (3), the physical causes all the mental effects. Finally, because of the Causal Exclusion Principle, premise (5), since there is a complete physical cause for everything, both mental and physical, the mental substance can cause nothing. Hence, *EASD* appears to entail Epiphenomenalism.

This argument fills one loophole left open by the Exclusion Argument Schema. We recognize that occasionally more than one complete cause occurs for a single effect. For example, two bullets may simultaneously strike a person's heart, and, thus, both bullets cause the person's death. Thus, in some rare cases, more than one complete,

independent cause brings about a single effect. So, *EASD* adds premise (6) that mental causation, if it exists, is not a case of overdetermination.

The problem of mental causation as presented here may seem unfamiliar. Traditionally, the problem of mental causation for Cartesian substance dualism has centered on the Physical Causation of the Mental and the Causal Closure of the Physical, premises (3) and (4) of *EASD*. However, the premises of *EASD*, with the exception of premise (2), are supported by good evidence.

The purpose of constructing an Exclusion Argument for Cartesian Substance Dualism (or, as I shall say interchangeably, Cartesian Dualism or Substance Dualism) is to show the generality of the Exclusion Argument, that it need not assume either Constitution Physicalism or Token-Physicalism. So, for the moment, I will simply assume Substance Dualism, and give some reason to accept the other premises.

2.2.1 Causal Closure of the Physical

Cartesian Substance Dualism is the view that minds and bodies are separate substances, distinct fundamental existents, that somehow causally interact with one another. The essence of the mind is thinking; the essence of body is extension. Mental substance is an immaterial stuff whose nature is to think or be conscious, whereas physical substance is a material stuff whose nature is to take up space. Philosophers have not improved significantly on the notion of an immaterial substance since Descartes' formulation, but philosophers have improved on the notion of the physical. The characterization of the physical as an extended substance is inadequate because some physical entities are considered to be unextended. For example, photons and other particles, clearly physical objects, are considered to be point masses, and thus

unextended. No single characteristic appears likely to provide the essence of the physical, but it is often assumed that the nature of the physical is captured by the sum of properties to which a completed basic physics would appeal.

This characterization, of course, does not mean that our physicists determine the nature of the physical, just that the entities are the sorts of entities to which any correct theory of physics will appeal. For this characterization of the physical to work, the putative physical theory must also be complete since one would not want to rule out forces or entities that have not yet been discovered. Assuming our current physics is largely correct, the physical could be characterized by a set of well-defined entities and their properties, like charge, mass, velocity, momentum, energy, etc.

There is good reason to suppose the Causal Closure of the Physical is true. Details about the substances aside, philosophers have demanded since Descartes' time to know how the mental, immaterial substance could affect the physical world. Descartes' interactive Substance Dualism, the view that these two substances causally affect each other, violates Causal Closure of the Physical, premise (4) of *EASD*, but such a violation seems impossible. Some philosophers have thought that it was inconceivable that two such different types of substance could interact. How could an immaterial mind have any effect on ponderous, physical matter? Other philosophers argued that interaction required violations of physical conservation laws, conservation of momentum and mass/energy. Descartes' critics in essence assumed something like the principle of the Causal Closure of the Physical.

Although apparently compelling, the conservation law objection to Cartesian Dualism assumes the Causal Closure of the Physical rather than supporting it. Laws of

conservation of momentum and mass/energy say that for any system closed to outside influence the total momentum and mass/energy of the system must remain constant. So, the objection that interaction requires a violation of conservation laws assumes that the physical is closed to outside, i.e. nonphysical, influence. The conservation-law objection thus assumes the Causal Closure of the Physical rather than giving evidence for that principle.

The conceivability objection against interaction is more convincing when interpreted correctly. Conceivability is not a guarantee of truth; nor is lack of conceivability a guarantee of falsity. However, it is hard to imagine how one could even begin to answer questions about how a mental substance could 'push' on the particles or alter the fields that make up the physical universe. The inconceivability of this interaction places the burden of proof on the Substance Dualist. The Substance Dualist must provide both a plausible and testable theory of the mechanism by means of which the immaterial substance could influence the physical and a convincing philosophical argument that makes such interaction intelligible. Without a theory of interaction including these features, we must suppose that the physical is closed to mental causal influence. Criticism of Cartesian Dualism fits fairly nicely the Causal Closure premise of the Exclusion Argument; it is inconceivable that the immaterial mind influence the physical body.

2.2.2 Physical Causation of the Mental

Although the Substance Dualist would, presumably, deny the Physical Causation of the Mental, premise (3) of the EASD, the evidence from neuroscience provides reason to accept this premise. The success of neuroscience in finding correlations between

mental events and neural events suggests that every mental event is caused (or otherwise determined) by the neurophysiology of the brain. Evidence of mental effects of brain lesions, the results of PET (Positron Emission Tomography) scans, the effect of chemicals, including chemicals as mundane as alcohol, all clearly support a dependence of the mind on the brain. Provided one accepts Substance Dualism, the Physical Causation of the Mental is well supported by the neuroscientific evidence.

This premise does not fit the traditional problem of mental causation for Descartes. The problem of how the physical could cause mental effects is historically less important but still significant. Premise (3) of *EASD*, the Physical Causation of the Mental, is the principle that all the mental effects that occur have a physical cause. The Cartesian picture denies this premise because it claims that only some mental events are determined by physical causes. According to Cartesian Dualism, one mental event can cause another without any intervening physical activity. However, the relevant problem with Substance Dualism is not that all the mental effects have physical causes, but how mental effects can have physical causes at all. Just as immaterial substance seemed incapable of affecting physical substance, ponderous physical matter seems incapable of affecting an immaterial mental substance. So, *EASD* appears to capture part of the problem of mental causation for Substance Dualism, but it does not capture the whole of the traditional problem.

For this reason, when I give a solution to the Exclusion problem in the final chapters of this work, that solution will not apply to Substance Dualism. I will argue that the Exclusion Principle does not exclude the mental from causal relevance, but Substance Dualism has significant other problems with interaction so that my solution will not help

explain interaction for Substance Dualism.

2.2.3 Causal Exclusion Principle

Kim (1987a) gives the basic argument for an Explanatory Exclusion Principle. His argument for the Explanatory Exclusion Principle depends on his claim that having more than one explanation creates an epistemic tension. Finding two or more explanations for a single phenomenon creates an epistemic tension, a need to discover that either one explanation is correct and the other is not, or that both explanations are related as incomplete parts of one explanation or in some other way.

More generally, we see that two purposes of explanation are simplification and unification. Having multiple explanations of a single phenomenon does not simplify or unify the explanation but complicates the explanation. Showing dependencies or relations among different explanations better unifies the explanation. Thus, the practice of explanation should involve reducing multiple explanations to a single explanation and showing how the multiple explanations depend on each other.

If explanatory realism is true, an objective relation must hold between our theories and the world. If that is so, then the purposes of simplification and unification are taken as likely to lead to true explanations. If those explanations are true, then something about the world must correspond, in some way, with those explanations. So, Kim's Explanatory Exclusion Principle corresponds to a metaphysical Exclusion Principle. For the purposes of *EASD*, we can take the Exclusion Principle to be causal exclusion, but the argument above shows that we should accept a more general Exclusion Principle, called the Generative/Determinative Exclusion Principle². I will discuss the Exclusion Principle again

² Kim appeals to this principle only in his unpublished David Ross Boyd lectures at the University of Oklahoma. He calls the principle the Generative/Determinative Exclusion Principle to allow for general

in later parts of this chapter and more critically in chapter eight.

Finally, we must admit that not all events should be excluded from causal relevance. I noted earlier that this argument filled one loophole left open by the Exclusion Argument Schema. In cases of causal overdetermination, more than one complete cause can occur for a single effect. For example, two bullets may simultaneously strike a person's heart, and, thus, both bullets cause the person's death. Or two singers can sing a high-C and shatter a glass, but neither one is the single cause of the shattering. Thus, in some rare cases, more than one complete, independent cause brings about a single effect. So, *EASD* adds premise (6) that mental causation, if it exists, is not a case of overdetermination.

This additional premise is justified as well as necessary for the argument. Cases of causal overdetermination are rare and accidental. They are a comparative oddity that cannot be explained by any appeal to a deeper underlying cause. However, the case of mental-physical causation is ubiquitous and systematic. Every mental event already has a physical cause, and assuming that the mental is an additional overdetermining cause does not fit the profile of overdetermination. The systematic overdetermination of mental events, having both a physical and a mental cause, strains the bounds of credibility. Such a systematic overdetermination cries out for further explanation, and yet to assume that such causation is simply a case of overdetermination gives up on further explanation. Overdetermination of the mental by both mental and physical causes seems to require us to believe in an incredible happy coincidence. So, we should not

determination relations including Goldman's (1970) notion of generation of actions from basic actions. Thus, on Goldman's view, my moving my arm quickly is not the same action as my moving my arm. Goldman argues that the basic action of moving my arm generates the less fundamental action of my moving it quickly. Kim wishes to include such generation relations with his principle.

assume that mental causation is a case of simple overdetermination.

So, *EASD* appears to be valid, at least on a cursory inspection, and the premises of that argument appear to be justified. The Physical Causation of the Mental is supported by evidence from the neurosciences, and it appears inconceivable that the Causal Closure of the Physical be false. Therefore, *EASD* shows that the mental, if Substance Dualism is correct, must be epiphenomenal.

Because Descartes denied both Physical Causation of the Mental and Causal Closure of the Physical, both premises (3) and (4) of *EASD*, the traditional problem of mental causation for Substance Dualism is not generally captured by the Exclusion Argument. Nonetheless, it is important to note that an Exclusion Argument can be constructed for Substance Dualism. If the premises of the Exclusion Argument as given above are true, then the mental substance must be causally inert or epiphenomenal. If it is impossible for the mental to be epiphenomenal, as I will argue, then Substance Dualism is inconsistent with the broadly physicalist premises of the Exclusion Argument. It is hardly surprising that Substance Dualism is inconsistent with these broadly physicalist premises, but, despite the differences between the traditional Cartesian problem of mental causation and the problem as given by the Exclusion Argument, a version of the Exclusion Argument applies to Substance Dualism.

Since the Exclusion Argument is not the traditional problem of mental causation, giving the Exclusion Argument in terms of Substance Dualism may seem irrelevant. In addition, few philosophers now accept Substance Dualism, so no one is apt to defend this view. Nonetheless, that the Exclusion Argument can be given against Substance Dualism shows that the Exclusion Argument is not strictly speaking a problem for Physicalism.

Since an Exclusion Argument can be constructed for Substance Dualism, philosophers like Kim (1993b) and Van Gulick (1992 and 1993) who include physical monism or Token Physicalism as essential to the Exclusion Argument are mistaken.

The only way for Token Physicalism to be necessary is for the thesis to be equivalent to a premise that is essential. Without getting too far ahead of ourselves, we can see that the only way that Token Physicalism can play an essential role in the Exclusion Argument is if it guarantees either that all the causation is physical or that everything is determined by the physical. Substituting Token Physicalism for the Causal Closure requirement will guarantee that all causes are physical, but it will guarantee nothing about the causal relevance of properties or types. Furthermore, substituting Token Physicalism for the determination requirement would guarantee that everything be determined by the physical if everything is physical, but it will guarantee nothing about determination of properties or types. Thus, Token Physicalism cannot be essential to all versions of the Exclusion Argument. Indeed, a solution to this problem of mental causation might be to abandon Substance Dualism in favor of some variety of physicalism. So, Token Physicalism seems more likely to be a solution to the problem of mental causation than part of the problem.

Although some broadly physicalist assumptions are essential to any version of the Exclusion Argument, the claim that everything is physical is not one of them. Because of the differences between the traditional problem of mental causation for Substance Dualism, and because of the current unpopularity of Substance Dualism, I will pursue *EASD* no further but will focus on versions of the Exclusion Argument for Nonreductive Physicalism. I will now turn to these versions of the Exclusion Argument.

2.3 The Exclusion Argument and Constitution Physicalism

I have shown that a version of the Exclusion Argument can apply to Substance Dualism, so this problem of mental causation is not strictly a physicalist problem. Other versions of the Exclusion Argument apply to different varieties of Physicalism. The second version of the Exclusion Argument applies to event dualism and concludes that mental events cannot be causes. The third version of the Exclusion Argument applies to Token Physicalism combined with Property or Type Dualism.

An increasingly common view of the mind and other macroscopic entities, called *Constitution Physicalism* or *Composition Physicalism*, is that the mind is constituted by or composed of only physical entities but is neither type nor token-identical to them. It is not possible to defend this view in detail, but I can briefly give one argument for the view. In brief that argument is that macroscopic objects have different modal properties from the sums of their parts, and so cannot be identical to sums of those parts. Gibbard's (1975) example of Goliath and Lumpl, a statue and the clay that constitutes it, shows that macroscopic objects and their constituent matter differ in modal properties³. Goliath could exist without his left thumb, for example, but, since Lumpl is the lump of clay that constitutes Goliath, Lumpl could not exist without that piece of clay. On the other hand, Lumpl could survive being smashed into a shapeless blob, whereas Goliath could not survive being smashed in this way. Assuming that names designate the same object in all possible worlds, the names "Goliath" and "Lumpl" cannot designate the same individual since one might exist without the other existing. Thus, the argument goes, macroscopic entities in general are neither type nor token identical to their constituent matter.

³ Gibbard's conclusion is not mine. He concludes that identity of objects designated rigidly is contingent, not necessary.

In what follows, I will assume for the sake of convenience that all causal relata are events, but if other entities, such as states or objects, can be causal relata, then the Exclusion Argument given here will apply *mutatis mutandis* to those entities as well. This assumption only applies, of course, if those other entities are constituted by physical parts but are not themselves identical to the sums of those parts.

Exclusion Argument for Constitution Physicalism (*EACP*)

- (8) *Event dualism*. Mental and physical events are real events.
- (9) *Irreducibility*. Mental events are not token-identical to physical events.
- (10) *Mereological Supervenience*. The mental is determined by the physical. All mental events supervene on physical events. Any two events with all their physical parts indiscernible in an indiscernible physical structure will be mentally indiscernible. Fixing all the parts and their relations to each other will fix the whole.
- (11) *Causal Closure of the Physical*. There can be no cause of a physical event that is not itself a physical event.
- (12) *Causal Exclusion Principle*. There can be no more than one complete, independent cause for any event. If there is one complete, independent cause for an event, there cannot be another cause for that event, except in cases of overdetermination.
- (13) Mental causation is not a case of causal overdetermination.
- (14) Therefore, *Token-Epiphenomenalism*. Mental events cannot cause any effects.

Constitution physicalism is in one sense dualist and in another sense physicalist. It is dualist in that it involves entities that are not even token-identical to physical entities, but it is generally classified as physicalism because macroscopic or mental objects are wholes made up entirely of physical parts. The view claims the only existents are physical entities and their aggregates. If no physical entities existed, then nothing in the universe would exist. To borrow Kripke's (1972) image, if God created everything physical and put it in its place, then he would not need to do any additional work or create anything else. If Physicalism is the view that nothing exists apart from physical entities and their aggregates, then Constitution Physicalism is a kind of Physicalism.

However, since Physicalism may reasonably be interpreted as the view that nothing exists apart from physical entities, Constitution Physicalism is also a kind of dualism.

Terminology need not concern us. The important fact for the purposes of the Exclusion Argument is that Constitution Physicalism does involve irreducibly distinct sets of entities. This fact is sufficient for the Exclusion Argument to apply to Constitution Physicalism.

The Exclusion Argument presented here should now be familiar. Since the mental and physical are distinct real existents, and the physical is closed to nonphysical causes, the mental can cause nothing physical. For example, on this view, my pain is constituted but is not identical to some underlying physical event, say a C-fiber firing. I'll assume that neural firings are physical events. The C-fiber firing causes another neural event, p_2 , which underlies a mental event, m_2 , my belief that I am in pain. Given the Causal Closure of the Physical, my pain cannot cause the neural event p_2 . And since the physical determines the mental, the mental need play no role in determining the mental. My belief that I am in pain is determined by the neural event p_2 , so my pain has no causal role to play. Since the mental and physical are determined completely by the physical, and assuming the Causal Exclusion Principle, the mental is causally irrelevant. My pain cannot cause my belief that I am in pain.

The Exclusion Argument presented here, however, is invalid. My talk of the physical determining both the mental and the physical masks a distinction that needs to be made for the argument to be valid. Physical events are caused by other physical events, but mental events are not directly caused by physical events. Mental events are determined by their constituent physical parts but are not *causally* determined by them.

The determination of mental events by physical events is not causal but of some other type, perhaps ontological. Wholes depend for their existence on their parts, but this dependence is not causal. Intuitively, without suggesting a counterfactual analysis of the relation, the whole could not exist without its parts. Physical events, then, may have only physical causes, but mental events do not appear to have physical causes. This fact may allow for the possibility that mental events cause other mental events.

For this Exclusion Argument to be valid, one needs to appeal to a broader Exclusion Principle. Kim's Generative/Determinative Exclusion Principle, mentioned earlier, will suffice.

Generative/Determinative Exclusion Principle: there can be no more than one complete, independent determiner of any event; if there is a complete, independent determiner of an event, then there is no other determiner of that event.

The reasons for accepting this view have been discussed already. Assuming that events are explananda, they require an explanans. If there is both a mental cause and a physical determiner of an event, then the event has two distinct explanations. But these two distinct explanations create an epistemic tension. The two explanations cannot be complete, independent and correct.

2.3.1 Causal Closure of the Physical

The reasons for accepting the Causal Closure principle are essentially the same for *EACP* as they were for *EASD*. It is hard to conceive how macroscopic objects could affect their physical events except through the physical events that constitute them. How could my pains affect my neurons except by the neurons that constitute my pain causing other neurons to fire? There does not seem to be a way for the mental to affect the physical without some physical implementing mechanism.

2.3.2 Mereological Supervenience

One reason the Exclusion Argument is seen as more naturally an argument against Physicalism is that the problems of interaction between different substances do not seem as insurmountable. Constitution Physicalism does not provide a means independent of the physical for the mental to cause physical events. However, Mereological Supervenience is much clearer than the Physical Causation of the Mental assumed by *EASD*. There did not seem to be any way that brute physical matter could have any effect on the mental substance. However, it is more plausible on a physicalist view that the physical determines the mental. The notion analogous to Physical Causation of the Mental, Mereological Supervenience, is clear and plausible.

Mereological supervenience is the thesis that properties of wholes, macroscopic objects and possibly minds, depend on and are determined by the properties of their parts and the relations among those parts. If the microconstituents of a macroscopic object or event are fixed, and the physical relations among them are fixed, then the properties of the macroscopic object are fixed. Somewhat more precisely, we can state weak Mereological Supervenience as:

Necessarily, for any object x , that is constituted by the mereological sum y , defined as proper parts a_1 through a_n in relations R_1 through R_n , and for some property F in A , if y has F , then x has some property G in B ⁴.

Weak supervenience states that within a world whenever the subvenient objects or supervenience base objects have a certain property, then the whole will have a

⁴ This characterization is potentially inaccurate because, as many philosophers argue, mereological sums may not themselves be objects or events. Use of the term y for the mereological sum should be taken as a convenience and not as a claim with ontological import. For this reason I eschew Kim's (1987b) formulations since they require there to be two objects.

It is important to note that the necessity here need only be nomic necessity. Because of the nature of the evidence for these claims, it is not safe to assume a necessity stronger than nomic. Fortunately, since causation is no stronger than a nomic relation itself, the type of necessity in the supervenience claim need

correlative property. So, for example, if my neurons have a set of properties, say my C-fibers are firing, then I will have a corresponding mental property, pain. And any time a person has her C-fibers fire, she will feel pain. However, weak supervenience is compatible with the possibility that I feel an itch when my C-fibers fire, so long as everyone else in the world feels an itch when their C-fibers fire. An example from a moral context should illustrate this point (from Kim 1984a). Imagine Socrates has the nonmoral properties of honesty, wisdom, etc. Let us assume that Socrates also has the moral property *goodness*. According to weak supervenience anyone else who has these nonmoral properties must also be good. However, Socrates in another possible world could have these nonmoral properties and be evil, according to weak supervenience. Everyone else in that world who had those nonmoral properties would also be evil. Weak supervenience says nothing about what is essential to the being with those properties. According to weak supervenience, goodness is not essential to anyone who has those nonmoral properties. So, weak supervenience seems too weak to capture our commonsense notion of dependence of moral properties on nonmoral properties. Moreover, weak supervenience seems too weak to capture the dependence of macroscopic objects on their parts. For example, in another possible world, my chair could be made up of exactly the same parts that it has in the actual world, but my chair in that other possible world could have less mass than it actually has. Thus, weak supervenience is too weak to capture our notion of dependence.

One can state a principle, Strong Mereological Supervenience, that better captures the notion of dependence:

Necessarily for any object *x*, that is constituted by the mereological sum *y*,

only be nomic.

defined as proper parts a_1 through a_i in relations R_1 through R_i , and for some property F in A , if y has F , then necessarily x has some property G in B

Strong supervenience does not allow for the problematic variation allowed by weak supervenience. According to strong supervenience, if Socrates has the properties of honesty, wisdom, etc., then he *must* be good. If my chair has all the parts and relations among them that it actually has, then it *must* have the mass it does. So strong supervenience, in particular strong Mereological Supervenience, better captures the notion of dependence we have in mind.

Three types of evidence support Strong Mereological Supervenience. Basic intuitions support the principle for ordinary properties. Successes in reductive explanation support the principle. Finally, evidence from the neurosciences support the principle for mental states in general, and even for conscious states for which we do not know and perhaps never will have adequate explanations.

First, the principle is intuitively almost undeniable for common examples of macroscopic properties. Ordinary properties of objects, the mass, rigidity, solidity, etc. of a chair, seem to be obviously dependent on the parts that make up the chair and their organization. So intuition provides some reason to accept the principle for some ordinary properties.

Second, Mereological Supervenience is a precondition for much of the reductive success of the natural sciences; it is the metaphysical assumption that we grounds our scientific strategies of understanding wholes by understanding their parts and the relations among those parts. Reductive explanation, taken as explanations of macroscopic objects in terms of their parts, in science is nearly ubiquitous and undeniable. For example, natural science explains temperature in a gas as mean

molecular kinetic energy. Solidity of objects is explained as electromagnetic fields in the molecular structure of the objects repelling each other. Science provides examples too numerous to mention of macroscopic objects and properties that can be explained in terms of their parts and the relations among those parts. Thus, success of the natural sciences supports Mereological Supervenience as a general assumption of science.

Third, successes in neuroscience suggest that this principle is true at least for correlations of mental states with neural states. Even conscious states appear to correlate in the required way. Philosophers⁵ have argued that we can never completely explain consciousness itself or the correlations between the physical and conscious states, but the existence of such correlations is well supported. For example, my conscious experience of a red image in my visual field may not be explicable in terms of the neural processes of the visual system, but there appear to be well-established correlations between particular color sensations and certain triplets of spiking frequencies in the visual system⁶. So, even in the most problematic case of consciousness, there is reason to think the covariation of Mereological Supervenience appears to obtain. Thus, the *prima facie* evidence for Mereological Supervenience is substantial. I will discuss supervenience and possible failures of supervenience in more detail in chapter seven.

Given the premises of the Exclusion Argument, with the addition of the Generative/Determinative Exclusion Principle, mental events cannot be causes. If, as I will argue in Part II of this work, chapters four through six, the mental is not epiphenomenal, then at least one of the premises of this version of the Exclusion Argument must be false or it must be invalid. One might decide that the problem of

⁵ Especially Nagel (1974) and Levine (1983).

⁶ The example of color as a triple of spiking frequencies is suggested by Churchland (1989).

mental causation could be avoided by rejecting the first premise of *EACP*. If mental events are identical to physical events, then it is possible for mental events to be causes because physical events are causes. Nonreductive Token Physicalism, however, has its own version of the Exclusion Argument with which to contend.

2.4 The Exclusion Argument and Nonreductive Token Physicalism

The third version of the Exclusion Argument entails a rather different form of Epiphenomenalism than the previous two versions. The third version raises the problem for Nonreductive Token Physicalism, the view that mental event tokens are identical to physical event tokens. Thus, any particular mental entity is identical to some particular physical entity. However, the Exclusion Argument requires that type physicalism, that each mental type is identical to a physical type, be false.

Although I cannot argue in detail for Nonreductive Token Physicalism, I can provide a brief argument against Type Physicalism. It has become philosophical conventional wisdom in recent years that the mental is multiply realizable in heterogeneously disjunct physical states. Pains may be realized in C-fiber firings in human beings, by some other brain state by octopi, and in some state of a silicon-based system in the case of putative Martians. These systems appear to have nothing *physical* in common. Therefore, there is no physical type that is identical to the mental type *pain*. Thus, an event's mental properties are distinct from its physical properties.

The Exclusion Argument for Nonreductive Token Physicalism is as follows. The mental and the physical are distinct properties of event tokens. When a mental event causes some event, the physical properties of the cause are all that is necessary for the physical effect to occur. The physical properties of any event determine the mental properties of that

event. Therefore, the physical properties of the cause determine the mental properties of the effect. Hence, the physical properties are causally sufficient for the effect. Since there can be no more than one complete, independent property causally relevant to any effect, mental properties are causally irrelevant. Any effect occurs only in virtue of the cause falling under a physical type and not in virtue of its falling under a mental type.

Exclusion Argument for Nonreductive Token Physicalism (*EANTP*)

(15) *Property dualism*. Mental and physical properties are real.

(16) *Irreducibility*. Mental properties are not identical to physical properties. For an event to fall under a mental type is not the same as for the event to fall under a physical type.

(17) *Property Supervenience*. The mental is determined by the physical. Any objects that are indiscernible with respect to their physical properties must be indiscernible with respect to their mental properties.

(18) *Causal Closure of the Physical*. There can be no property causally relevant to the production of a physical effect that is not itself a physical property.

(19) *Generative/Determinative Exclusion Principle*. There can be no more than one complete, independent property causally relevant for the production of an effect. If there is one complete, independent property causally relevant for the production of an effect, there cannot be another property causally relevant for the production of that effect, except in cases of overdetermination.

(20) Mental causation is not a case of causal overdetermination.

(21) Therefore, *Type-Epiphenomenalism*. Mental properties cannot be causally relevant to the production of any effects.

2.4.1 Token Physicalism and Event Identity

I need to discuss two influential views of events, the fine-grained and coarse-grained conceptions of events, and how accepting one or the other might affect this third variety of the Exclusion Argument. The fine-grained conception of events appears to collapse the distinction between the second and third versions of the Exclusion Argument. The fine-grained view of events, Kim's (1976) view, is that an event is a complex consisting of an object exemplifying a property at a time. Thus, my turning on the light switch at 10:00 p.m. consists of me exemplifying a property. *turning on the light switch* at the time of 10:00 p.m.

On this view of events, my startling the burglar is not the same event as my turning on the light switch because each event exemplifies a different property. Since mental and physical properties are distinct, for familiar reasons, mental events are not identical to physical events. So, on the fine-grained view of events, there is no distinction between *EACP* and *EANTP*.

The coarse-grained view of events, Davidson's (1969) view, takes events to be simple, fundamental constituents of the world. My turning on the light switch and startling the burglar can be one and the same event with two different descriptions or with two different properties. Mental events, on this view, can be identical with physical events even if mental and physical properties are not identical. Obviously, it would be simpler for purposes of exposition to assume the coarse-grained view of events. However, I do not wish to argue the relative merits of the two conceptions of events, and I do not want to beg important questions by assuming either view. Instead, I will present the Exclusion Argument in a form that will apply to either notion of events.

The fine-grained view of events is incompatible with Token Physicalism conceived as the view that all events are physical events. The only way for the fine-grained view of events to be compatible with Token Physicalism is for mental properties to be identical with physical properties. Since irreducibility of properties is one of the assumptions of Nonreductive Physicalism, it follows that for the fine-grained view of events Nonreductive Physicalism is automatically event dualism.

However, the event-dualism that derives from this conception of events is not a substantial dualism. One might not be able to avoid a dualism about events on this conception, but one could still be a physicalist about the constituent objects. The objects that

exemplify certain properties may be identical, and so the view itself may be a form of identity theory. Since, on the fine-grained view, events are not the basic constituents of the world, the objects that in part constitute events are the fundamental entities that should be the relata in the claims of identity theorists. Property dualism may entail event dualism on a fine-grained conception of events, but, since events are not fundamental constituents of the world, property dualism does not entail dualism about fundamental entities. One could be an identity theorist about objects without committing oneself to any dualism deeper than property dualism.

Constitution Physicalism, conversely, accepts a substantial dualism about events beyond anything forced on one by a particular conception of events. Let us assume for a moment the fine-grained conception of events. For Constitution Physicalism the objects that in part make up events are not identical to the mereological sums of their parts. On this view, mental events are not identical to physical events for two reasons, because the objects exemplifying the properties are not identical and, possibly, because the properties exemplified are not identical. Constitution Physicalism entails dualism about events even if the coarse-grained analysis of events is correct. The coarse-grained notion of events by itself is compatible with mental events being either type or token-identical to physical events. But Constitution Physicalism entails that both type and token physicalism are false. So, Constitution Physicalism entails a more substantial dualism about events than Token Physicalism does.

Perhaps the problem of mental causation is a reason to accept one conception of events over another since the the fine-grained conception appears to collapse an important distinction. However, I think that even assuming that conception of events, an

important distinction can be made for causal relevance. If events are not fundamental constituents of the universe, and they are causal relata, then it follows that the objects, properties and times that constitute them must be causally relevant. On Constitution Physicalism, macroscopic objects are not identical (either type or token) to sums of their constituent objects. So, even if we accept a fine-grained conception of events, we can still distinguish questions about the causal relevance of the objects constituting events and the causal relevance of the properties constituting the events. Thus, the Exclusion Argument for Constitution Physicalism would claim, at least, that macroscopic objects and mental objects are not causally relevant; and the Exclusion Argument for Nonreductive Token Physicalism would claim only that properties of mental events are not causally relevant. Thus, even on a fine-grained conception of events, we can deduce a distinct variety of the Exclusion Argument for each dualism. Because of this fact, I will assume a coarse-grained conception of events for ease of exposition.

2.4.2 *EANTP* and *Quausation*

EANTP addresses the properties of physical events and their causal relevance rather than addressing the events and whether they can be causes. Following the terminology introduced by McLaughlin, whereas the first two versions of the argument lead to Token-Epiphenomenalism, the third version of the Exclusion Argument leads to Type-Epiphenomenalism.

Horgan's (1989) neologism *quausation*, mentioned in chapter one, will help in understanding this Exclusion Argument. Horgan characterizes this notion as follows.

For any two events *c* and *e* and any two properties *F* and *G*, *c qua F* causes *e qua G* iff:

- (i) *c* causes *e*;
- (ii) *c* instantiates *F*;

- (iii) **e instantiates G; and**
- (iv) **the fact that c instantiates F is explanatorily relevant to the fact that e occurs and instantiates G. (1989, 50)**

On Horgan's notion of quausation an effects occurs and is the type of event it is in virtue of the type of the cause. For example, some event token in my brain may instantiate a pain type and a C-fiber firing type. That is to say, one event is both a pain and a C-fiber firing. This event causes me to pull my hand from the hot stove. To say that the C-fiber firing quauses the event of my pulling my hand from the hot stove is to say that my pulling my hand away occurs in virtue of that cause being a C-fiber firing. Quausal relevance seems an integral part of explanations of events and of why those events are of the type they are. Again, assuming explanatory realism, there must be some objective fact that makes these explanations correct. Thus, mental quausation must occur iff behavioral and mental effects, occur in virtue of the mental property of the cause. And *EANTP* entails, at least apparently, that the mental quauses nothing. According to *EANTP* Type-E or Quausal Epiphenomenalism, as Horgan terms it, is correct.

2.4.3 Generative/Determinative Exclusion Principle

EANTP differs from the *EACP* in that it assumes properties, or the fact that an event instantiates a type, are explananda. If we assumed only a Causal Exclusion principle or the Generative/Determinative Exclusion Principle, it is logically possible that the mental properties of the cause be relevant to the mental property of the effect. The mental properties of any event are determined by the physical properties of the event but are not caused by them. Thus, the Causal Exclusion principle does not exclude the possibility that the mental property of a cause is relevant to the mental property of the effect. The Causal Exclusion principle allows the mental property to quause another

mental property. The G/D Exclusion Principle, in addition, simply does not address exclusion of properties of events but only of the events themselves. Thus, for the Exclusion Argument to be valid, we must formulate Kim's G/D Exclusion Principle to include properties.

Generative/Determinative Exclusion Principle for properties: there can be no more than one complete, independent determining set of properties for any property (or the fact that an event *e* instantiates a property); if there is a complete, independent determining set of properties for a property (or the fact that an event *e* instantiates a property), then there is no other property that determines that property (or the fact that an event *e* instantiates a property).

According to this principle, if the mental properties of an effect are determined, as Property Supervenience says they are, then there can be no other determiner of the mental property. Thus, this principle excludes the possibility that a mental property of a cause may quause the mental property of the effect. The G/D Exclusion Principle for properties rules out the possibility of mental quausation.

This Exclusion Principle is justified in the same manner as the others. As long as the fact that some property occurs can be explained, then an Explanatory Exclusion principle should apply to it. And if explanatory realism is true for properties, then this principle is as well justified as any previous Exclusion Principle. The fact that my experience now is a pain seems in as much need of explanation and as susceptible to explanation as the occurrence of the experience itself. Therefore, it seems obvious that property instantiations require and are susceptible to explanation.

2.4.3 Causal Closure of the Physical

The principle of the Causal Closure of the Physical (*CCP*) also needs to be amended to deal with properties. The first formulation for *CCP for properties* is as

follows.

Causal Closure of the Physical for properties: No non-physical property can be relevant to any physical effect; for any caused event *e* that has a physical property *P*, no non-physical property of *e*'s cause can be relevant to the occurrence of *e*.

This principle asserts that no mental property can be relevant to the occurrence of a physical effect if no mental property is a physical property. However, it makes no claims about the physical properties of that effect. Thus, the first formulation is sufficient for the conclusion that no non-physical property is relevant to a physical occurrence, but it does not entail anything about any possible physical property of the effect. For example, this principle would exclude the possibility that my pain might cause some purely physical event, say a brain firing *b* correlated with my belief that I am in pain. However, it does not exclude the possibility that the fact that the cause of *b* is a pain is relevant to *b*'s being the neural state it is.

Assuming this version of the *CCP*, we cannot be certain that the mental properties of the cause do not determine the physical properties of the effect. So, the first formulation does not make *EANTP* valid. The second formulation of the *CCP* for properties remedies this problem. It is as follows.

Causal Closure of the Physical for properties: No non-physical property can be relevant to any physical effect or for the fact that the physical property is instantiated; for any caused event *e* that has a physical property *P*, no non-physical property of *e*'s cause can be relevant to the occurrence of *e* or to the fact that *e* instantiates *P*.

This principle seems as safe as the previous Causal Closure principles. It is hard to imagine how a mental or otherwise nonphysical property could be relevant to a physical effect or to the fact that an effect is physical without some implementing physical property. How could my pain be relevant to the neurons in my brain firing or to

the neurons in my arm firing without some neural basis? Thus, Causal Closure of the Physical properties seems secure.

2.4.4 Property Supervenience

Finally, *EANTP* requires Property Supervenience to entail Type-Epiphenomenalism. Assuming supervenience, the physical properties of the effect determine its mental properties. Following Kim (1987b), we formulate strong supervenience between A properties, or subvenient properties, and B properties or supervenient properties, as follows.

Necessarily, for any object x and any property F in A, if x has F, then there exists a property G in B such that x has G, and *necessarily* if any y has G it has F. (1987b, reprinted 1993a, 80)

The physical properties, the A properties, of an object determine the mental properties, the B properties, of that object. The reasons for accepting strong supervenience are approximately the same as for accepting Physical Causation of the Mental and Mereological Supervenience. Evidence from the neurosciences suggests a close correlation between brain states and mental states; whenever one is in a particular brain state, it is nomically necessary that one be in a particular mental state. Thus, whenever my C-fibers fire, I must feel pain. The supervenience appears to be of this strong variety; C-fiber firings *must* correlate with pains. I will discuss the varieties of supervenience necessary for the Exclusion Argument in more detail in chapter three.

2.5 Conclusion

I have shown that a version of the Exclusion Argument infects any type of dualism given a set of plausibly true, broadly physicalist premises. One can construct an Exclusion Argument for Substance Dualism. This argument does not capture the traditional problems

of mental causation for Substance Dualism since the greatest problem for Substance Dualism is the inconceivability of interaction with the physical. However, the fact that an Exclusion Argument can be constructed for Substance Dualism shows that the problem is not one limited to physicalism, strictly conceived, but for any dualism that includes certain broadly physicalist premises.

Exclusion Arguments can be constructed for both Constitution Physicalism and Nonreductive Token Physicalism. The Exclusion Argument presents perhaps the most serious problem for these two views. Although there are differences in the premises of these latter two Exclusion Arguments, the similarities between them will be more important than their differences. Although it is possible that one variant or isomorph of a premise for one Exclusion Argument can be true, and the isomorphic premise for another Exclusion Argument be false, the differences between isomorphic premises will not make a difference to which theory best deals with the problem of mental causation. I will note different commitments of the variants of the Exclusion Argument when it becomes necessary later in this work. The most important fact to note is that isomorphic Exclusion Arguments can be given for both types of Nonreductive Physicalism.

The remainder of this work will address the Exclusion Argument for these two varieties of physicalism. The next chapter of this work will show the relations between the premises presented in this chapter and other alternative premises and will give reasons to prefer the premises I have chosen for the Exclusion Argument.

CHAPTER THREE

VARIANT PREMISES OF THE EXCLUSION ARGUMENT

3.1 Introduction

Philosophers have presented the Exclusion Argument with premises distinct from those I have used thus far. In this chapter, I will explain the entailments between sets of premises assumed by two different versions of the Exclusion Argument. In particular, I will discuss Exclusion Arguments that assume Supervenience of Causal Powers and arguments that assume Property Supervenience and Causal Closure of the Physical. I will show that the premise of the Property Supervenience together with the claim of the Causal Closure of the Physical entails the claim of Supervenience of Causal Powers; I will show that the premise of Supervenience of Causal Powers with some uncontroversial assumptions entails the principle of the Causal Closure of the Physical. I will then discuss various forms of the principle of Causal Closure of the Physical, and show which version is most plausibly true. Finally, I will suggest that a version of the Exclusion Argument that assumes Property Supervenience and Causal Closure of the Physical is superior to a version that assumes supervenience of causal powers.

3.2 Supervenience of the Mental on the Physical

Property Supervenience of the mental on the physical is roughly the claim that, for any object, if it has a particular physical property, then it has a particular mental property. Philosophers put this claim in several ways. One way is that indiscernibility with respect to the physical entails indiscernibility with respect to the mental. Another way of putting this claim is that there can be no difference in the mental properties of an object without some difference in its physical properties although it leaves open the possibility that there may be a

physical difference in the object without there being a mental difference. For example, for some event, if it is a C-fiber firing, that event is a pain. Thus, supervenience allows for multiple realizability of mental properties in different physical properties but not for mental properties to vary when the physical properties are fixed¹.

Supervenience has been thought to be an analysis of dependence, the claim that one set of properties or facts depends on another set of properties or facts. At least one way for some item, A, to depend for its existence on some other item, B, is for B to cause A, but there are more forms of dependence than causal dependence. Another form of dependence is ontological dependence. Ontological dependence is a relation between items some of which are more fundamental than others and that the less fundamental items require the more fundamental for their existence. This way of expressing the idea of dependence suggests a supervenience account.

Recently, however, philosophers have argued that supervenience is inadequate as an analysis of dependence. (See Grimes, 1988 and Kim, 1990) The criticisms have focused on the fact that supervenience only involves covariation (or necessary covariation) of properties, and thus is too weak a notion to capture the ontological force of a dependence claim. Necessary common determinants provide counterexamples to the account of dependence as supervenience since supervenience would be unable to distinguish the determinant from the necessary common byproducts. Schematically, two properties A and B may be determined by a necessary common determinant C. A supervenes on B, but B does not supervene on A.

¹ As a simplifying assumption, I will only discuss supervenience of properties of the same object, not supervenience of properties of wholes on the properties of their parts, when the wholes are not identical to the sums of their parts. I believe that everything I have to say in this chapter about the one type of supervenience will hold for the other type as well.

In this case, A would supervene on B but not depend on B but on C. Kim, giving a concrete example, writes

I've heard that there is a correlation between intelligence as measured by the IQ test and manual dexterity. It is possible that both manual dexterity and intelligence depend on certain genetic and developmental factors, and that intelligence strongly covaries with manual dexterity but not conversely. If such were the case, we would not consider intelligence to be dependent on, or determined by, manual dexterity. (Kim, 1990, repr. 1993a, 146)

The possibility of a case like that described by Kim shows that dependence cannot be captured by supervenience because supervenience, even asymmetric supervenience, only involves a (necessary) covariation of properties. Since this criticism contends only that supervenience is insufficient for dependence, it will be safe for the purposes of the Exclusion Argument to assume (as Kim does) that supervenience constitutes a necessary condition of dependence. Thus, supervenience is adequate as an assumption of the Exclusion Argument even if it does not capture the ontological force required of dependence.

Kim (1987b) formulates three different types of Property Supervenience. For generality Kim puts the relation in terms of A properties, the supervenient properties or, for our discussion, the mental properties, and B properties, the subvenient or base properties, for our discussion, the physical properties.

Weak supervenience:

Necessarily, for any object x and any property F in A, if x has F, then there exists a property G in B such that x has G, and if any y has G it has F. (1987b, reprinted 1993a, 80)

Strong supervenience:

Necessarily, for any object x and any property F in A, if x has F, then there exists a property G in B such that x has G, and necessarily if any y has G it has F. (1987b, reprinted 1993, 80)

Global supervenience:

Any two worlds indiscernible with respect to B-properties are indiscernible with respect to A-properties. (1987b, reprinted 1993a, 82)

As Kim argues, global supervenience is too weak to capture ordinary notions of dependence because it allows for difference of mental properties without a significant physical difference. To take an example like Kim's, global supervenience is consistent with the existence of a world identical to the actual world except that it contains one fewer hydrogen atom on Jupiter and in that world my C-fiber firings correspond not to pains but to contemplation of mathematics or to no conscious mental state at all. So, if only global supervenience is true, it is possible that mental properties may not correlate with brain properties when only a clearly irrelevant difference occurs across possible worlds. Thus, global supervenience is too weak to capture our ordinary notion of dependence.

In addition, global supervenience guarantees nothing even within a possible world about the relation between local physical properties and mental properties. Specifically, it is possible that any object that has a particular physical property, say being a C-fiber firing, might be a pain or it might not be a pain. If only global supervenience obtains, my C-fiber firing may correlate with pain and yours may correlate with contemplation of mathematical truths. Or, worse, my C-fiber firing at time t_1 and my C-fiber firing at t_2 may not correlate with pains. As long as fixing all the physical properties of the world fixes all the mental properties in that world, global supervenience is true. However, global supervenience allows for my C-fiber firing at t_1 to correlate with pain whereas my C-fiber firing at t_2 and your C-fiber firing at any time correlate with contemplation of mathematical truths. As long as every world with exactly the same physical characteristics has these mental characteristics, global supervenience obtains. Thus, global supervenience does not guarantee that I will feel pain when my C-fibers fire or guarantee anything else about individuals' mental and physical

states. So, because global supervenience guarantees nothing across possible worlds and nothing about individuals within a possible world, global supervenience is too weak to capture dependence.

Global supervenience may, nonetheless, be the strongest type of supervenience compatible with externalism about mental content. Mental content, as Burge (1979) and Putnam (1975) argue, is not determined solely by local physical properties. For example, since I know nothing about beeches and elms except that they are trees, my thought that beeches are trees and my thought that elms are trees do not differ in local properties, either psychological or physical. However, the two thoughts differ in mental content because they are beliefs about different objects. Examples like these make a strong case that mental content does not supervene on local physical properties.

I will return to this issue in detail in chapter seven, but for now it is important to note that global supervenience is not strong enough to capture anything like our notion of dependence or for the Exclusion Argument to be valid. For the Exclusion Argument to be valid, physical properties of individual objects must fix the mental properties of those objects.

Weak supervenience, on the other hand, may capture a notion of dependence and is strong enough to make the Exclusion Argument valid. According to weak supervenience, necessarily, for any object that has a particular physical property, it will have a particular mental property. Fixing the physical properties fixes the mental properties within a world. It may seem that global and weak supervenience are equivalent since neither requires a difference in mental properties across possible worlds, at least possible worlds in which there is any physical difference. However, weak supervenience guarantees that individual objects

with the same physical property will have the same mental property at least within a possible world, whereas global supervenience guarantees nothing about individual objects. So it must be the case that my C-fiber firings at different times are all pains if the property of being a pain weakly supervenes on the property of being a C-fiber firing.

Strong supervenience may be the best assumption for the Exclusion Argument. Strong supervenience gets the closest to dependence of the three types of supervenience, and so it is likely the closest to the actual relation between the mental and physical. If only global or weak supervenience of the mental on the physical obtains, then the mental properties of an object could differ across possible worlds without any difference in that object's local physical properties. In addition, the Exclusion Argument is apparently valid (worries about the Exclusion Principle aside) with strong supervenience as the supervenience assumption.

Assume for the moment the mental supervenes only weakly on the physical. If causal relevance has no modal force, then the Exclusion Argument is valid. In that case, the mental cannot be causally relevant if the mental is fixed by the physical within a possible world. Therefore, the Exclusion Argument is valid assuming only weak supervenience if causal relevance has no modal force.

However, if causal relevance is a modal notion, then weak supervenience may not be enough to make the Exclusion Argument valid. The argument for this claim is as follows. Assume causal relevance has some modal force. Thus, if the physical properties of an event are causally relevant to its effect, then the physical properties make the effect necessary. If that is the case, then the physical property of an event e_1 could determine the physical property for its effect e_2 across possible worlds, but if only weak supervenience holds, then

the mental cannot be fixed across possible worlds by the physical. If the mental is only fixed by physical properties in one world, and causal relevance holds across worlds, then there may be more to mental causal relevance than weak supervenience can capture. Any Exclusion Argument that assumes only weak supervenience may allow the mental property to be causally relevant by determining a mental property of an effect across possible worlds.

The mental could still be causally relevant if only weak supervenience obtains. Consider the case of two events e_1 and e_2 . In possible world w_1 , e_1 has two properties, a physical property P_1 and a mental property M_1 . Assuming Causal Closure of the Physical for properties, P_1 , and not M_1 , is causally relevant to e_2 's having some physical property P_2 . Assuming that causal relevance involves some necessity, e_1 's having P_1 necessitates e_2 's having P_2 . In world w_1 , if the mental weakly supervenes on the physical, when e_2 occurs, and has P_2 , e_2 also has some mental property M_2 . If e_2 has P_2 in possible world w_2 , but only weak supervenience obtains, then e_2 may have some other mental property M_3 in w_2 . Any effect may have one mental property in world w_1 and a different mental property in w_2 if the mental only weakly supervenes on the physical. The physical property of a cause in one possible world cannot fix the mental properties of an effect across possible worlds if the mental only weakly supervenes on the physical. So it is possible that M_1 might necessitate that e_2 have M_2 in w_2 . It is possible that the mental might be causally relevant by fixing mental properties across possible worlds if the mental only weakly supervenes on the physical.

Relying on weak supervenience to make room for mental causation would mean that no difference could be discerned in the actual world, so the causal relevance of the mental would never be apparent. Thus, we might doubt whether the causal relevance allowed for by

weak supervenience really exists. This type of skepticism about modal properties was common in early twentieth-century metaphysics, but it is not justified. Similar reasoning would show, for example, that causation does not exist because it cannot be discerned from mere conjunctions in the actual world. This reasoning really counts against any modal claims at all, so as long as we have any modal intuitions, we should be willing to admit causal relevance as having a modal dimension.

Accepting only weak supervenience and rejecting strong supervenience is not a promising way of making room for the causal relevance of the mental. General skepticism about modality and essential properties provide the only reason to accept weak supervenience over strong supervenience. This skepticism should apply to the modality of causal relevance exactly if it applies to supervenience. Strong supervenience supposes a *de re* necessity just as I am supposing causal relevance does. So the potential solution of accepting only weak supervenience of the mental on the physical but assuming modal force of causal relevance is an uneasy solution at best. The reasons for accepting or rejecting *de re* modal claims are the same in either case.

Furthermore, there is good reason to accept strong supervenience as obtaining for dependence relations. Kim (1984a) gives the following example to suggest to bring out intuitions that some *de re* modality is inherent in a notion of dependence. He writes,

The idea of strong supervenience comes to this: if St. Francis is a good man, there must be some combination of these virtues (say, honesty and benevolence) such that St. Francis, and anyone who has it *must* be a good man. (Kim, 1984, repr. 1993, 65)

Strong supervenience fits our intuition that these virtues make it necessary that St. Francis is good. If only weak supervenience obtained for moral characteristics, if St. Francis had those qualities, then he might not be a good man. In the actual world, if he had those

characteristics, he would be a good man, but he might not be good even with the same nonmoral characteristics. Current philosophy of mind supposes that the mind similarly depends on the brain. A wide variety of evidence, from studies of brain lesions and their effect on cognitive, emotional and conscious states, to first hand evidence of the effects of alcohol and drugs on the mind, shows that the mind depends on the brain. If we accept this evidence that the mental depends on the physical, we should think that, say, when my C-fibers fire I must feel pain. Given the occurrence of that event, it has to be the case that I feel pain. So, if we share these modal intuitions, we should accept that strong supervenience obtains for cases of dependence, specifically the case of the dependence of the mental on the physical.

Therefore, I suggest that accepting strong supervenience as a working hypothesis is justified by our modal intuitions. We can safely assume strong supervenience as the type of supervenience assumed by the Exclusion Argument. Anyone lacking these modal intuitions may wish to accept only weak supervenience, but this is unlikely to provide much relief from the problem of mental causation because such a philosopher should not agree that causal relevance has modal force. The Exclusion Argument could, then, be put in terms of weak supervenience only along with the assumption that causal relevance has no modal force. However, since most philosophers accept the truth of some *de re* modal intuitions, I will assume strong supervenience of the mental on the physical.

Ultimately, I believe, failure of both strong and weak supervenience does not allow for mental causation, but it is the burden of chapter seven of this work to show that claim. For the moment, I shall be concerned with the validity of the Exclusion Argument and making a *prima facie* case for the premises necessary for the argument.

3.3 Commitments of Supervenience

I need to note some important facts about the ontological commitments of supervenience. First, although there are other formulations of the supervenience thesis that avoid explicit appeal to properties, accepting the reality of the mental and physical requires some form of dualism. For example, one might accept Hellman and Thompson's (1975) model theoretic account of supervenience in order to avoid a commitment to the existence of properties. However, to suppose that their account is true is to commit this version of supervenience to some form of dualism. If, for example, some fact about the world makes a model theoretic account true, then both the mental and the physical would be real. Presumably one could question whether the mental and physical, as real entities, were causally relevant to an effect. And so, one could generate an Exclusion Argument for whatever commitments one's formulation of supervenience requires, so long as one is ultimately a realist about the mental and physical.

However, the supervenience thesis itself does not entail realism about either the F or G sets of properties, the mental or physical properties. Supervenience does not assume that either the F or G set of properties exists. It does not even assume that the properties are possible. The property of being a round square supervenes on the property of being a square, but the property of being a round square is not even a possible property. Supervenience merely is the relation that in all possible worlds, if an object and a property from one set of properties exist, and that object has that property, then there will be a property from another set of properties which that object also has. So, the supervenient and the subvenient properties may be nonactual. If a supervenient property exists, then so does a subvenient property, but unless there is such a supervenient property, there need not be a subvenient

property. This observation is relevant to the mind-body problem in that, for example, an Eliminative Materialist could agree that mental properties supervene on physical properties, but still claim that there are no mental properties. Therefore, we cannot dispense with the premise of dualism since supervenience does not entail it. The supervenience thesis alone does not commit us to the existence of the properties in question.

We should also note that we need to assume the realism of both the physical as well. It is more fashionable in recent years to doubt the realism of the mental, but anti-realism about the physical could also defeat the Exclusion Argument. If, for example, the physical were merely instrumental or a convenient fiction, then one would not think that the physical was genuinely causally relevant. So, on this phenomenalist or instrumentalist view, the mental may have all the genuine causal relevance, and the physical, as a mere fiction, would have none. Although it may be that no philosopher currently holds to phenomenism or instrumentalism about the physical, the validity of the Exclusion Argument depends on physical realism just as it does on mental realism.

The final note is that supervenience does not entail distinctness of the F and G sets of properties, or, in our case, the mental and the physical. One way to see this fact is to see that, consistently with supervenience, the F and G properties may be identical. Supervenience is neither a symmetric nor an asymmetric relation². Supervenience relations may be one-way, like the putative supervenience of the mental on the physical or the supervenience of the temperature on the mean kinetic energy of a gas. In these cases the supervenient property is

² Symmetry and asymmetry as formal properties of a relation R must hold for any instance of that relation R. So, the relation of *being to-the-immediate-left-of* is asymmetric since when A is immediately to the left of B, B cannot be immediately to the left of A. The relation of *being-adjacent-to* is symmetric since whenever A is adjacent to B, B is also adjacent to A. But some relations, like *being a brother*, are neither symmetric nor asymmetric, since A can be a brother to B, and B can be a brother to A if B is male, but if B is female, B is not a brother to A. Supervenience is like the *brother-of* relation since we can stipulate it to

multiply realizable in disjunctive physical properties, so the subvenient properties do not also supervene on the physical. For example, temperature may be realized in a gas, solid, liquid or as background radiation in a vacuum. Fixing temperature, therefore, does not fix the realizing, or subvenient, property, but fixing the underlying physical property fixes the temperature. In the case of asymmetric supervenience the sets of properties are distinct.

However, supervenience does not guarantee distinctness. Since supervenience may also be two-way, identical properties are cases over which supervenience obtains. Take for example, some property Q. Trivially, fixing the property Q of an object fixes the property Q of that object. And any property Q essential to an object strongly supervenes on the essential Q of that object. Or, put another way, if properties F and G are identical, then trivially any x that has F must have G and vice versa. In fact, a type identity theorist would believe the mental supervenes on the physical. According to the type identity theory, mental properties are type identical to physical properties, so that mental language and physical language are merely different descriptions of the same property. This view is compatible with strong supervenience since indiscernibility of the physical would entail indiscernibility of the mental for the case in which mental and physical properties are identical. Indeed strong supervenience is a necessary condition for and a straightforward consequence of the type identity theory.

Since the type identity theory avoids the problem of mental causation by denying that mental and physical properties are irreducible, and not by denying supervenience, for the Exclusion Argument to be valid, assuming only strong supervenience of properties is not enough for the Exclusion Argument to be valid. The Exclusion Argument requires separate

be symmetric if A and B properties are identical (or nomically equivalent) or asymmetric if the relation between A and B properties only necessitates in one direction.

assumptions of realism of the mental and the physical and irreducibility (or failure of type identity) of the mental to the physical. Assuming only supervenience does not guarantee these additional claims. Thus, I have shown why the Exclusion Argument needs the premises which I have assumed, and supported, to be valid. I will now discuss some variant premises that one might use in giving the Exclusion Argument.

3.4 Entailments among Variant Premises

Philosophers have given versions of the Exclusion Argument that differ from the Exclusion Argument Schema as presented in this work. Fodor (1989) for example, takes as a premise the Supervenience of Causal powers. Baker (1993) takes as premises Property Supervenience and Causal Closure of the Physical. If these arguments are different versions of the same argument, then it should be possible to show that the assumptions are equivalent. Or, if they are not equivalent, one could discover which argument makes the least controversial assumptions necessary for the Exclusion Argument to be valid. In that case, one should find that one set of premises entails the other but not vice versa. If few or none of the premises entail each other, then it might be best to consider them different arguments for the same conclusion.

The thesis of Property Supervenience does not alone entail the thesis of Supervenience of Causal Powers (as one might think) but requires another principle, the Causal Closure of the Physical. The thesis of Supervenience of Causal Powers does entail the thesis of Causal Closure of the Physical but does not entail the thesis of Property Supervenience.

First, I should clarify the notion of a causal power. The term ‘causal power’ appears fairly frequently in the literature³ without much discussion of its nature. Causal powers could be either intrinsic (non-relational) or relational properties of an object or event in virtue of which that object or event has the effect that it does. To illustrate the difference between these views, I will borrow Block’s (1990) example of the matador’s cape and the bull. On the first view, that causal powers are intrinsic properties of objects, the redness of the matador’s cape is the property that is causally relevant to enraging the bull. On this account of causal powers, objects or events have causal powers but properties do not have causal powers. On this view, causally relevant properties do not *have* causal powers, they *are* causal powers. The redness of the cape is the power of the cape to enrage bulls.

On the second account, causal powers are second-order properties or properties of properties. For example, the redness of the matador’s cape has the causal power or capacity to enrage bulls, but the redness is not itself the causal power to enrage bulls. Thus, we could say that properties have causal powers but not that objects or events have causal powers except derivatively by having properties that have causal powers.

Kim (1993b) appears to take a more pluralist approach to the analysis of causal powers. In discussing, what he calls Alexander’s Dictum, “To be real is to have causal power.” (202) He writes “I believe this principle, as applied to concrete *existents* and their *properties*, will be accepted by most non-reductive physicalists.” (202, italics added) Kim here thinks of causal powers as properties of properties and as properties of objects. On this view a property could have a causal power but that property need not *be* a causal power.

Ultimately, on either account, causal powers are properties of objects. Second-order properties are ultimately properties of the objects because properties only have causal power

³ Especially in Fodor (1987) and (1989).

when they are instantiated. Abstract particulars, particular instantiated properties, have causal power, but uninstantiated universals cannot have causal power. In fact, the lack of causal power of abstract entities is the primary reason some philosophers have for skepticism about them. So, causal powers, as second order properties of objects, are ultimately properties of objects, and, as such, will be covered by the supervenience theses above.⁴

Since causal powers are properties of objects, the theses of Property Supervenience and Causal Closure of the Physical entail the thesis of Supervenience of Causal Powers. The reverse is not the case. The thesis of Supervenience of Causal Powers does not entail the thesis of Property Supervenience, but it does entail the principle of Causal Closure of the Physical. These relations, although I have not yet proven them, show that if one rejects the thesis of Supervenience of Causal Powers, then one must also reject the thesis of Property Supervenience. Although one could reject the thesis of claim of Property Supervenience without rejecting the narrower Supervenience of Causal Powers, the reasons for rejecting the thesis of Property Supervenience will in general be reasons for rejecting the thesis of Supervenience of Causal Powers. Finally, if the physical is not causally closed, then mental causal powers do not supervene on physical causal powers although the physical could be causally closed without mental causal powers supervening on physical causal powers.

3.4.1 Causal Powers

The theses of Property Supervenience and Causal Closure of the Physical entail the thesis of Supervenience of Causal Powers. We have already seen Kim's formulation of the supervenience of one set of properties on another. We can formulate the supervenience of causal powers in a similar manner.

⁴ Some analyses of causal powers, especially Fodor's (1989) invoke causal laws, but these laws are not themselves properties or causal powers of objects, so we can safely ignore this issue for now.

Strong Supervenience of Causal Powers:

Necessarily, for any object x and any causal power F^* in A^* , if x has F^* , then there exists a causal power G^* in B^* such that x has G^* , and necessarily if any y has G^* it has F^* .

The argument that strong supervenience of properties entails that mental causal powers supervene on physical properties begins as follows.

- (1) Necessarily, for any object x and any property F in A , if x has F , then there exists a property G in B such that x has G , and necessarily if any y has G , it has F .**
- (2) Therefore, necessarily, for any object x and any causal power F^* in A^* , if x has F^* , then there exists a property G in B such that x has G , and necessarily if any y has G , it has F^* .**

The conclusion (2) appears to follow immediately from (1) by an instantiation of the universal in the strong Property Supervenience claim. However, the inference from (1) to (2) requires more argument, and that argument will differ depending on what theory one has of causal powers.

There seem to be three ways to view the nature of mental causal powers. One view, the view with the greatest intuitive plausibility, is that mental causal powers are identical to (perhaps a subset of) mental properties. On this view, this subset of mental properties supervenes on physical properties, provided the Property Supervenience thesis is true. If causal powers are properties of objects, then there is be no problem in making the inference from (1) to (2). For example, if mental causal powers are identical to mental properties, pain is identical to the causal power an entity has to engage in avoidance behavior. Thus, we can validly instantiate some causal power F^* for the property in A in the strong Property Supervenience claim, and arrive at claim (2). Thus, on this view mental causal powers supervene on physical properties.

The second possibility is Block's (1990) view that mental properties are identical to second-order properties of physical properties. Block conceives mental properties as second-order properties ranging over first-order physical (or at least lower-level) properties. So, on Block's view, mental causal powers are properties of physical properties, so, given Block's theory of the nature of the mental as second-order properties, the mental properties and the mental causal powers are identical. Thus, on this view, again, fixing the physical properties fixes those mental properties. The A set of properties, the mental properties, are causal powers, so these causal powers supervene on physical properties if the Property Supervenience thesis is true. So, on Block's view of causal powers, the inference from (1) to (2) is valid.

The third possible view is that mental causal powers are properties of mental properties, then the simple inference from (1) to (2) above is not valid. Causal powers, on this view, need not be properties in the A set of properties. For the case of the mental and the physical, it is possible that the mental causal powers are not mental properties at all but properties of those mental properties. Thus, mental causal powers need not be mental properties. For example, the causal power of my pain, say its capacity to cause me to engage in avoidance behavior, may not itself be a mental property. The causal powers mental properties *have* may not be themselves mental properties.

An appeal to transitivity of supervenience solves this problem. The view that presents a problem to the apparently simple inference from (1) to (2) is the claim that causal powers are themselves properties of the properties in the A set; that is, the mental's causal powers are not, on this view, mental properties at all but are in some other set of properties entirely. This view, which makes our simple inference invalid, still assumes Property

Supervenience thesis, and so the causal powers of mental properties themselves supervene on the physical properties by transitivity of supervenience. The mental causal powers, as properties of mental properties, supervene on the mental properties themselves. Because of the causal specification of second-order properties, second-order properties supervene on their first-order realizing properties. If my pain causes me to act necessarily in certain ways, then my pain meets the causal specifications of the second-order property, the causal power my pain has. Since that is the case, the physical properties still fix mental causal powers.

So we can add an additional premise, for this view of causal powers, that the mental's causal powers supervene on the mental properties. Premise (4) makes this claim for some causal power H in the set of second-order properties C.

- (3) Necessarily, for any object x and any property F in A, if x has F, then there exists a property G in B such that x has G, and necessarily if any y has G it has F.
- (4) Necessarily, for any object x and any property H in C, if x has H, then there exists a property F in A such that x has F, and necessarily if any y has F it has H.
- (5) Therefore, necessarily, for any object x and any causal power H in C, if x has H, then there exists a property G in B such that x has G, and necessarily if any y has G it has F*.

The conclusion (5) claims that fixing the physical properties, the property G in B, fixes the mental causal powers, the property H in C. Since fixing the physical properties fixes the mental properties, and fixing the mental properties fixes the mental causal powers, fixing the physical properties also fixes the mental causal powers. These three views exhaust the conventional theories of causal powers as second-order properties, so we should suppose that mental causal powers entails that mental causal powers supervene on physical properties.

It follows from the Property Supervenience thesis, on any of the three conventional views of causal powers, that mental causal powers supervene on physical properties. However, because the consequents of the component conditionals in (1) and (2) are existential claims, we cannot also assume that any property in B that we instantiate to must also be a causal power. Although the mental causal power supervenes on a physical property, it does not follow that the mental causal power supervenes on a physical causal power. Since properties are a broader category than causal powers, if Epiphenomenalism is possible, the mental need not supervene on only those properties that are causal powers.

If causal powers are properties, in particular some subset of the properties of an object or event, then the thesis of Supervenience of Causal Powers appears to be a simple entailment of the Property Supervenience claim. However, it does not follow from the supervenience of mental properties on physical properties and that causal powers are properties of objects, that mental causal powers supervene on physical causal powers. If causal powers are a subset of the properties of an object or event, then mental causal powers must supervene on physical *properties* but not necessarily on physical *causal powers*. Nothing in the claim of Property Supervenience entails that the mental causal powers will supervene only on the physical causal powers. These claims leave open the possibility that at least some non-causal physical properties are included in the set of properties the mental causal powers supervene on. Hence the argument above does not establish the principle of strong Supervenience of Causal Powers.

We could avoid this difficulty by assuming that all properties have or are causal powers. Kim calls this Alexander's dictum (Kim 1993b) after an early twentieth century emergentist Samuel Alexander who made this assumption in his (1927). Kim writes, "*To be*

real is to have causal powers." (202, Kim's italics) For a concrete particular or its properties to exist at all it must have causal power. Using the notion of causal powers given earlier, all causal powers are properties. Assuming Alexander's dictum in the context of the Exclusion Argument would simply beg the question. If we wish to allow, at least for the moment, that Epiphenomenalism is a coherent possibility, we cannot as yet conclude that all properties are causal powers. We wish to allow for the possibility of properties that are not and do not have causal powers. Therefore, we need some further argument to connect the two theses.

3.4.2 Causal Closure of the Physical

This fact shows why thesis of Property Supervenience requires the Causal Closure of the Physical principle to entail the Supervenience of Causal Powers. The mental causal power might, if the physical is not causally closed, supervene on some noncausal physical property instead of on a causal power. Thus, unless the Causal Closure of the Physical principle or some similar claim is true, Property Supervenience could allow for the mental to have an effect on the physical directly even if the mental is determined by some physical property. If only Property Supervenience obtains, it is possible that the physical property not be causally relevant while the supervenient mental property is.

Before continuing, I would like to digress briefly on the thesis of the Causal Closure of the Physical. A more formal characterization of the causal closure of the physical than we made use of in the previous chapter roughly follows Kim (1995). One way to characterize the causal closure of the physical as a general claim about properties in some set of properties *D* as follows.

Necessarily, for any object or event *x*, if *x* has a cause, there exists some *F'*, in some set of properties *D*, that is causally relevant to the occurrence of *x*.

This formulation differs from our less formal varieties in the previous chapters in two important ways. First, it is nonexclusive. This formulation allows that a property outside the D set of properties could be causally relevant to the occurrence of x as long as some property in D was also causally relevant. Second, this formulation does not make the Exclusion Argument valid. This formulation has the flaw, discussed in the previous chapter, that it does not require that the physical properties, but only the events, be fixed only by the physical. Thus it would allow for downward causal relevance for the occurrence of physical properties. And if the physical properties of the effect are not fixed, then the mental properties of the effects are not fixed either. Thus, this statement of Causal Closure does not require that the physical properties of a cause fix the physical or mental properties of the effect.

This version of the Closure principle, we discovered, was not sufficient to make the Exclusion Argument valid. We needed a closure principle for the physical properties not just for effects with physical properties.

Thus, there are four important versions of the causal closure principle. Two of them are exclusive as follows.

Exclusive Causal Closure of the Physical:

Necessarily no non-physical event causes a physical event.

Exclusive Causal Closure of the Physical for properties (or Quausal Closure):

Necessarily no non-physical property is causally relevant to any physical effect or to any effect insofar as that effect is physical.

Nonexclusive Causal Closure Principles could be given as follows.

Nonexclusive Causal Closure of the Physical:

Necessarily, for any physical effect, there is a physical cause for that event.

Nonexclusive Causal Closure of the Physical for properties (or Quausal Closure):

For any physical property of an effect, necessarily, there is a physical property of the cause that is relevant to the physical effect's occurrence and for the physical effect's having that physical property.

It is possible to formulate a stronger type of causal closure, a *de re* modal claim that any particular effect has a particular cause essentially. The weaker Causal Closure principles merely claim that there must be some physical cause for any effect, but they do not specify a particular cause that must accompany a particular effect. For example, say my arm movement away from a hot stove is caused by my C-fiber firing. On the *de dicto* closure claims, the arm movement must have some physical cause, but that movement may be caused by some different brain firing than it actually is caused by. A *de re* closure claim would not allow for this. The *de re* principle may be given as follows.

Strong causal closure:

Necessarily, for any effect, there is an event that necessarily causes that effect.

There is good reason, however, to reject the strong causal closure principle. It claims that one cause or causally relevant property is essential to an effect, but it seems unlikely that this claim is true. An event could have many different causes, or perhaps no cause at all. Intuitions about causes suggest that we reject the strong Causal Closure principle and its variants.

Moreover, the Exclusion Argument does not require a strong Causal Closure of the Physical. All that is required to rule out mental causes is that all the causes be physical (and not mental), and not that any particular physical effect have a particular physical cause. Thus, the causal closure principle that should be accepted should be one of the *de dicto*

principles given above, depending primarily on which type of physicalism one accepts. The Causal Closure of the Physical for properties is necessary for Nonreductive Token Physicalism, whereas the simple Causal Closure of the Physical is necessary for Constitution Physicalism.

Whether to accept an Exclusive or Nonexclusive Causal Closure principle is a more difficult question. The Nonexclusive Causal Closure principle leaves open the possibility that both a mental and physical property may be causally relevant to the production of a physical effect, whereas the Exclusive principle does not allow for the mental to be causally relevant to the physical. Because the Exclusion Argument assumes an Exclusion Principle, it is valid with either an Exclusive or Nonexclusive principle, so for the moment I will assume a Nonexclusive principle. The Generative/Determinative Exclusion principle when applied to causes of physical events can rule out the possibility that the mental is causally relevant to physical effects since, according to the Nonexclusive Causal Closure principle, the physical event already has a cause. The same can be said, *mutatis mutandis*, for physical properties. Finally, the argument for causal closure of the physical, that it is inconceivable that a physical effect might be caused by a mental event without some intervening or implementing physical property, only supports the Nonexclusive Causal Closure of the Physical principle. Thus, I will assume a Nonexclusive Causal Closure principle. I will discuss these issues at somewhat greater length in chapter eight. I will now return to the main argument of this chapter.

3.4.3 Thesis of Supervenience of Properties and Causal Closure of the Physical entail thesis of Supervenience of Causal Powers

The entailment from the claim about the supervenience of mental causal powers on physical properties to the claim that mental causal powers supervene on physical causal

powers requires a further premise to the effect that the physical causal powers are the only properties on which the mental causal powers can supervene. Without accepting that all properties are causal powers, we need to assume the additional premise of the Causal Closure of the Physical. One way to guarantee that the mental causal powers of an object supervene only on the physical causal powers and not on some non-causal properties is to say that the physical is closed to external causal powers. If a mental property supervened on some non-causal property and still was causally relevant to some physical effect, this occurrence would violate the Causal Closure principle.

If the physical is causally closed, mental causal powers supervene only on physical causal powers. According to causal closure for physical properties, no physical effect can occur in virtue of some mental property (or causal power) but not in virtue of some physical property (or causal power). In other words, if the Nonexclusive Causal Closure principle is true, mental causal powers cannot produce a physical effect without some physical causal power producing that effect. Because mental causal powers are properties, they supervene on physical properties. Mental causal powers cannot be causally relevant to the physical without some physical causal power being relevant to the physical effect. So, because of the Causal Closure of the Physical, mental causal powers supervene only on physical causal powers. Thus, the theses of Property Supervenience and the Causal Closure of the Physical entail the thesis of Supervenience of Causal Powers.

3.4.4 Thesis of Supervenience of Causal Powers entails thesis of Causal Closure of the Physical

The thesis of Supervenience of Causal Powers entails the thesis of Nonexclusive Causal Closure of the Physical. The claim of Supervenience of Causal Powers implies that the physical causal powers are enough for any effects that occur. This thesis, again, is the

claim that if one fixes the physical causal powers, then one fixes the mental causal powers. Indiscernibility of physical causal powers necessitates indiscernibility of mental causal powers. If supervenience of causal powers is true, every mental causal power is responsible for an effect only if there is some physical causal power correlated with that mental causal power. For any effect, there must be a physical causal power responsible for it. This conclusion is essentially the Nonexclusive version of the Causal Closure of the Physical.

3.4.5 Thesis of Supervenience of Causal Powers does not entail thesis of Property Supervenience

The thesis of Supervenience of Causal Powers does not entail the thesis of Property Supervenience. We have already noted that we should not yet assume that all properties are causal powers because this assumption immediately rules out the possibility of Epiphenomenalism. Supervenience of Causal Powers says nothing about the relations among putative non-causal properties. The thesis of Supervenience of Causal Powers would not entail that any non-causal properties that exist would have any relation to physical properties. Supervenience of causal powers makes no claim about any such non-causal properties. If there are some properties that are not causally relevant, then the thesis of the supervenience of causal powers has nothing to say about these properties. If these non-causal properties exist, they need not supervene on the physical properties. Therefore, the thesis of Supervenience of Causal Powers does not entail the thesis of Property Supervenience.

Since Fodor's version requires this somewhat weaker thesis, one might think that it is the better version. However, this difference is relatively slight and I think we are better off using the slightly stronger pair of theses because they allow for a more perspicuous set of premises. If we reject the causal supervenience thesis, then we may do it for one of two

reasons. We may reject the supervenience thesis in general or the Causal Closure of the Physical. The Exclusion Argument is clearer with these two premises instead of the single premise of supervenience of causal powers because the two premises allow us to differentiate two possible claims either of which of the two premises might be problematic.

If we want to solve the problem of mental causation with as little violence as possible to our metaphysical world view, we should use the version of the argument that allows us to reject only one premise and still keep all the others. Some philosophers attempt to solve the problem of mental causation by rejecting Property Supervenience without rejecting Causal Closure of the Physical. I am not yet concerned with whether these attempts are possible, but assuming the thesis of Supervenience of Causal Powers as fundamental to the Exclusion Argument would not allow one to make this distinction.

In any event I do not think the difference between these two versions of the argument will turn out to be relevant for our purposes. If we choose to reject the Supervenience of Causal Powers thesis, then we are necessarily rejecting the Property Supervenience thesis. Any reason for rejecting the thesis of Property Supervenience is likely to be reason for rejecting the slightly weaker thesis of Supervenience of Causal Powers. If we reject the Supervenience of Causal Powers thesis, we could not accept the general Property Supervenience claim, but we could still believe that the putative non-causal properties do not supervene on any physical properties. However, it is difficult to see why we would want to hold to that non-causal properties fail to supervene whereas causal powers do supervene. Any reason to reject the Supervenience of Causal Powers thesis is likely to be reason to reject the claim of supervenience of non-causal properties as well. I will simply note that the Property Supervenience claim applies to putative non-causal properties and the

Supervenience of Causal Powers claim says nothing about this set of properties. This slight distinction in the two formulations of the argument will make no real difference, and so I will use the more explicit set of assumptions.

3.5 Conclusion

In this chapter I have shown that the two variations of the Exclusion Argument the version assuming Property Supervenience and Causal Closure of the Physical and the version assuming Supervenience of Causal Powers are nearly equivalent. The first set of premises entails the second set and includes an additional claim that any putative non-causal mental properties supervene on the physical. The second set of premises says nothing about these putative non-causal mental properties. This slight difference in the formulations will not matter. Thus, the only reason to choose one of these versions of the argument over the other is because of perspicuity of premises. I take the version that assumes Property Supervenience and Causal Closure of the Physical to be the more perspicuous since it gives us two premises that we might have different reasons to reject.

In this part of the work I have attempted to show that the Exclusion Argument as I have given it is the most perspicuous version of the argument. It includes all the premises that are necessary for each of two varieties of Nonreductive Physicalism. It includes two distinct premises, those of Property Supervenience and Causal Closure of the Physical, which are conjoined in other varieties of the Exclusion Argument. Thus, the Exclusion Argument Schema includes all the premises one needs, and distinguishes different premises that may be rejected or accepted for different reasons.

In the next part of this work, I will argue that the conclusion of the Exclusion Argument cannot be true, and so the Exclusion Argument must be unsound. In Part III, I

will consider two premises, the thesis of Property Supervenience and the Exclusion Principle, that may be false or, in the case of the Exclusion Principle, broad enough to allow the mental to be causally relevant. But detailed discussion of these premises must wait until I establish that Epiphenomenalism is not an adequate reaction to the Exclusion Argument.

PART II:
ARGUMENTS AGAINST EPIPHENOMENALISM

OVERVIEW

One reaction to the problem of mental causation, especially as it appears in the Exclusion Argument, is to suppose that the mental is causally inert or epiphenomenal. This view, Epiphenomenalism, has been the bogey man of much recent philosophy of mind. Epiphenomenalism is frequently treated not as a position on the mind-body problem, but as a constraint on theorizing about that problem. A view's leading to Epiphenomenalism is taken as sufficient reason to reject or revise that view.

Two varieties of Epiphenomenalism, Type Epiphenomenalism and Token Epiphenomenalism, correspond to the conclusions of the two varieties of Exclusion Argument discussed in previous chapters. I will argue that Token Epiphenomenalism can be effectively refuted. I will also conclude that although no knock-down argument refutes Type Epiphenomenalism, there is good reason to think it false and no good reason, aside from the Exclusion Argument itself, to think it true. Thus, in later chapters I will consider which thesis, or piece of reasoning, of the Exclusion Argument could best be rejected.

In arriving at my admittedly weak conclusion, I will discuss several arguments against Epiphenomenalism. First, I will briefly examine several common arguments against Epiphenomenalism and conclude that they are not effective. Second, I will give a longer defense of Epiphenomenalism against the charge of incoherence. Third, I will argue that problems of knowledge of nonconscious mental entities will show that Token Epiphenomenalism cannot be true. Fourth, I will argue that the possibility of nomic

correlations of the mental with the physical make problems of knowledge, reference and representation of our own qualia, including the possibility of skeptical hypotheses, less than decisive criticisms of Epiphenomenalism. Nonetheless, I will argue, fifth, that the burden of showing that these nomic correlations could ground the relations just mentioned cannot be met by Epiphenomenalism, and so we should reject this view.

A brief reminder of the two varieties of Epiphenomenalism is necessary before discussing these arguments. McLaughlin (1989) following Broad (1925) distinguishes and characterizes two types of Epiphenomenalism, Type Epiphenomenalism and Token Epiphenomenalism. The distinction, in McLaughlin's words, is as follows,

Type Epiphenomenalism (Type-E). (a) Events can be causes in virtue of falling under physical types, but (b) events cannot be causes in virtue of falling under mental types.

Token Epiphenomenalism (Token-E). (i) Physical events can cause mental events, but (ii) mental events cannot cause anything. (1989, 109-10)

The characterization of Token-E is self-explanatory. Type-E, on the other hand, requires some explanation. Type-E is the view that the mental *qua* mental is not causally relevant, that an event insofar as it is a mental event is not relevant to the production of an effect. Only the physical properties of an event are causally relevant to the occurrence of an effect; the mental properties are not.

Type-E does not entail Token-E because token mental events may be identical to token physical events. If token mental events are identical with token physical events, then they are causes if and only if the physical events are. However, one need not think on this account that the mental properties are relevant to producing the effect since the effects may be causes in virtue of falling under the physical event type but not in virtue of falling under the mental event type. In other words, the mental properties would be causally irrelevant if

the effect occurs because they are physical events and not because they are mental events. Thus, an argument that Token-E is false will not entail that Type-E is also false.

Token-E does, however, entail Type-E. If mental events cannot cause anything, then no mental event can be a cause in virtue of its mental properties. If the mental event is not a cause, then no property of the event is causally relevant to producing any effect. Type-E is the weaker view because it is entailed by but does not entail Token-E.¹ Thus, a reason to think Token-E is false is not necessarily a reason to think Type-E is false.

I will not specify the variety of Epiphenomenalism until the distinction becomes relevant to the argument under consideration. Thus, for generality, I will continue the discussion in terms of Epiphenomenalism without distinguishing the two types.

At this point I will briefly examine several common arguments for the causal efficacy of the mental. The following is a list of these arguments. I will discuss the first five briefly and spend considerably more time on the last two.

1. Argument from Ordinary Language or Common Practice
2. The No-Causation-at-All argument
3. Fodor's End-of-the-World argument
4. Davidson's Argument from Explanation
5. Argument from Simplicity
6. Argument from Incoherence of Epiphenomenalism
7. Argument from First-Person Knowledge

In chapter four I will discuss the first five arguments fairly briefly. Then I will discuss the sixth argument in more detail. I will find that none of these arguments convincingly refutes the possibility of Epiphenomenalism. In chapters five and six I will discuss the final

¹ Philosophers recently have been more concerned with Type-E. Campbell (1970), for example, embraces it. Much has been written on the question of whether Davidson (1970) is committed to such a view. For example, Sosa (1984), Honderich (1982), Stoutland (1985), Johnston, (1985), Kim (1984b) and (1989b), and Dretske (1989) claim that he is committed to such a view. McLaughlin (1989) and (1993) and Lepore and Loewer (1987) claim that he is not.

argument at length for different types of mental entities, for non-conscious mental entities in chapter five and for conscious entities in chapter six.

CHAPTER 4

BRIEF ARGUMENTS AGAINST EPIPHENOMENALISM

4.1 Introduction

The first argument appeals to ordinary language use of causal terms or to ordinary practice of causal ascriptions to show that Epiphenomenalism is false. The second argument is that if there are no mental or special science causes, then our paradigm cases of causation are not cases of causation at all. Thus, it is possible that causes disappear altogether or are replaced by something altogether different from causation as ordinarily conceived. A third argument, implied by Fodor's (1989), is that our knowledge that our minds are causes is better justified than any philosophical argument leading to a contrary conclusion. The fourth argument, which comes from Donald Davidson (1963), is that for mental ascriptions to provide real explanations, the mental must be causally relevant. The fifth argument is that Epiphenomenalism countenances unnecessary entities and should be rejected by an appeal to simplicity. The sixth argument is an argument for the incoherence of Epiphenomenalism given some commonplace assumptions about the mind. The seventh argument is that Epiphenomenalism conflicts with the fact of our knowledge of and reference to our own minds. I will devote the next two chapters to this last argument.

4.2 Argument from Ordinary Language or Common Practice

The first argument is based on our ordinary language use of causal terms. The argument proceeds as follows. We use mental causal language in ordinary life, in folk psychological explanations, and in scientific, including scientific psychological, discourse. If Epiphenomenalism is true, then all of these causal explanations of our behavior are incorrect. According to Epiphenomenalism, my hunger does not cause me to eat, nor do my

beliefs cause me to act. Instead microphysical interactions cause me to act as I do. In accepting the Exclusion Argument, we must think that insofar as the psychological is irreducible to the neurophysiological and the neurophysiological is irreducible to the microphysical only the basic microphysical, perhaps quantum mechanical, could cause my actions. However, since our causal terms have their reference fixed not by use in quantum mechanical contexts but in mental, folk psychological and special science contexts, it must be that our causal terms in fact refer to these macrophysical events and properties. Therefore, so the argument goes, these macrophysical and mental properties or events are causally relevant.

This argument can hardly be convincing by itself since ordinary practice is defeasible. For example, ordinary language and common practice assumes that the sun rises and sets, but science tells us that the earth rotates. More seriously, common practice has at one time or another made false assumptions about the motion of the Earth, celestial bodies, and the nature of the physical world. Scientific advances often show that common practice is wrong. We might consider common practice to provide justification until some reason overrides or defeats that justification, as astronomy has overridden our assumption that the sun revolves around the earth. However, one might suppose that the Exclusion Argument itself provides a reason to defeat the prescriptions of common practice. The second argument, however, makes appeals to the Exclusion Argument less attractive.

4.3 The No-Causation-at-All Argument

The second argument is an objection to the Exclusion Argument rather than simply an argument against Epiphenomenalism. The No-Causation-at-All argument attempts to show that the Exclusion Argument must be unsound because if it were sound, there would be

no causation, or no recognizable causation, at all. This result is thought to be absurd, so the Exclusion Argument must be unsound. This argument does not affect Epiphenomenalism as such, but the Exclusion Argument, and our present reasons for accepting Epiphenomenalism. If we had independent reasons for Epiphenomenalism, this argument would not affect them². It is, nonetheless, appropriate to discuss this argument here since it is directly involved in the possibility that the Exclusion Argument might defeat the presumption in favor of ordinary practice.

Again, the second argument is that if the Exclusion Argument is sound, no causation at all occurs or the little causation that occurs bears no resemblance to our common conception. As Baker (1993) suggests, if we find the Exclusion Argument to be reason to reject the causal relevance of the mental, we must suppose that few recognizable properties are causally relevant. The Exclusion Argument holds for any sets of properties distinct from more basic properties. So, if the Exclusion Argument is sound, the only potentially causally relevant properties must be those in the most basic science. If quantum mechanics, which involves interactions that are fundamentally either indeterministic or nonlocally deterministic, is the most basic science, then the most basic science involves at least some events for which our concept of causation does not apply. According to quantum theory, properties, like position and momentum of particles, are indeterminate until an observation is made. There can be determinate facts only if the particles somehow "know" that a measurement is occurring or will occur. In either case, the causation left by the exclusion argument is far stranger than our normal concept of causation. Therefore, the argument

² In addition, one could avoid a rejection of the Exclusion Argument along these lines by trying to show that the Exclusion Argument does not generalize. This is Kim's (1995, 1997 and 1998) strategy. I find Kim's arguments unconvincing since he assumes the problem arises primarily from the thesis of physical realization of second-order properties. In contrast, I hope to have shown that the problem arises from the

concludes, for our talk of causation to have any referent, it must refer to those paradigm cases of causation involved in the special sciences and in mental causation.

This reasoning still does not suffice for rejecting Epiphenomenalism. Even if our ordinary notion of causation lacks the referent we previously thought it had, it is possible that it actually refers to some set of events in the world. Perhaps we could alter our notion of causation, so that we applied the term 'causation' to a different relation, just as we could apply the term 'unicorn' to horses when we discovered that unicorns did not exist. But using a term to refer in that way does not make the horses unicorns, nor does it make epiphenomena causes. So, in the case of causation, perhaps common practice uses the term to apply in cases in which it should not, but that fact does not make minds and the special sciences causally relevant.

Whether we continue to apply the word 'causation' in these cases is of little interest. It is more interesting whether causes really exist, not whether we choose to call certain relations causation. No doubt interesting distinctions among epiphenomena could be captured by the term 'causation,' but I am here concerned with whether mental causation really occurs

4.4 Fodor's End-of-the-World Argument

A third argument in a similar vein, suggested by Fodor (1989), is a Moorean style of argument. He writes,

Whereas, 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. (Ellipsis in original, 1989, 77)

broader assumption of supervenience.

Fodor argues here that the causal relevance of the mind is so fundamental to our beliefs that it is essentially unrevisable. I am unsure if Fodor intends this view, but a similar argument could be made that mental causation so obviously occurs that we must reject any argument that leads to the conclusion that it does not occur. Any argument that leads to such a conclusion must be more suspect than the position it attacks. Fodor, apparently, thinks that nothing could be more obvious and more fundamental to our understanding of the world than our mental states cause our actions.

However, the apparent obviousness of a thesis is not conclusive evidence for it, and it is hard to imagine an empirical thesis that should not at least in principle be open to revision. That the mental appears obviously causally relevant is not enough to overwhelm any argument against it. It seemed obvious that the earth stood still while the sun and stars revolved around it; it seemed obvious the world consisted of solid physical objects, but now we know that the earth revolves around the sun and physical objects consist mostly of atoms and empty space or perhaps something even less fathomable. Our understanding of the physical universe is arguably as fundamental as our knowledge of our own minds as causes. These beliefs about the physical universe certainly must have seemed as unrevisable as our beliefs about ourselves and the causal power of our minds seems to Fodor. Thus, although Fodor's point gives a compelling reason to discover a flaw in the Exclusion Argument, until we can show how it is possible for the mental to be causally relevant, we cannot reject Epiphenomenalism.

4.5 Davidson's Argument from Explanation

The fourth argument against Epiphenomenalism is given by Davidson (1963). His target is the view that reasons explanations are not causal explanations. On this view,

reasons explanations, or rationalizing explanations, do not give causes of actions but instead show how actions are rational. This view is not quite the Epiphenomenalism I am discussing because it denies that reasons are the right sorts of things to be causes at all. Epiphenomenalism, as I understand the view, claims that mental phenomena are causally irrelevant. This view of rationalizing explanations is that reasons are not causally relevant or irrelevant; understanding reasons in this way is a category mistake. It would be equally misleading to say that the number 1 did not make the last out in the baseball game; numbers are not the kinds of things that play baseball. Nonetheless, Davidson's argument is relevant because it appears to show that any purported non-causal explanation cannot be correct. He argues that for reasons explanations to be adequate explanations, reasons must be causes. Child (1994) summarizes the argument as follows.

Reasons explain actions. But we have not yet explained an action if we have said only that *S* acted some way and had a reason to perform that action: 'for a person can have a reason for an action, and perform the action, yet this reason not be the reason why he did it. Central to the relation between a reason and an action it explains is the idea that the agent performed the action *because* he had the reason.'³ But what is the force of the 'because' which captures the link between reason and action it explains? We have two options. Either we say that the link between reason and action is *sui generis*, a basic relation; in which case there is no further mystery. Or we give some kind of analysis of the relation; and what could that be but a causal analysis? As Davidson says: 'One way we can explain an event is by placing it in the context of its cause; cause and effect form the sort of pattern that explains the effect, in a sense of "explain" that we understand as well as any. If reason and action illustrate a different pattern of explanation, that pattern must be identified.'⁴ And, Davidson thinks, there is no plausible candidate for the pattern of reason explanation other than the causal pattern: 'the best argument for a [causal] scheme . . . is that it alone promises to give an account of the "mysterious connection" between reasons and actions.'⁵ (1994, 91)

I cannot add anything to Child's analysis of Davidson's argument. The confirmed Epiphenomenalist could avoid Davidson's conclusion in one of three ways: by claiming that

³ Davidson (1963) 691

⁴ *ibid.* 692

reasons explanations are *sui generis*, explanations of their own unique type, by claiming that reasons explanations are not real explanations at all, or by claiming that they are some other familiar type of explanation. The first response is unappealing in the context of the Exclusion Argument. Epiphenomenalism, when viewed as a reaction to the problem of mental causation, must not appeal to mysteries that are as great as problem of mental causation itself. The idea that reasons explanations are *sui generis* essentially appeals to one mystery to explain another and represents no significant advance over the notion that somehow mental causation just mysteriously occurs.

The second response suggests we take Davidson's argument as showing, not that reasons are causes, but that for reasons explanations to be real explanations, reasons must be causes. This response to Davidson's argument, combined with the next argument, shows that an Epiphenomenalist cannot afford to reject the use of reasons explanations, even in principle. I will briefly discuss this argument before moving to the third possible Epiphenomenalist response, that reasons explanations are some other familiar variety of explanation.

4.6 Argument from Simplicity

The fifth argument against Epiphenomenalism is that it should be rejected for its unnecessary ontology. Since epiphenomena, according to the second response to Davidson's argument, can play no role in our explanations, they should be rejected on grounds of simplicity. This argument, combined with Davidson's, seems unassailable. Although it may still be possible for an entity that plays no explanatory role whatsoever to exist, maintaining a belief in it is indefensible. Simplicity is no guarantee of truth, but belief

⁵ *ibid.* 693

in an explanatorily irrelevant entity is indefensible unless that entity is known independently of any explanatory role it may play.

The Epiphenomenalist can give two responses to these combined arguments short of rejecting Ockham's Razor. First, the Epiphenomenalist could respond that some causally irrelevant entities can still play a familiar role in explanations; this response is to repeat the third possible response to Davidson's argument. I will show in a moment what that familiar explanatory role might be. Second, the Epiphenomenalist should point out that we do not believe in the mental, at least our conscious mental states, because of some explanatory role they fill but because they are themselves our conscious mental life. Ockham's razor can only be used to reject explanations of phenomena, not to rejecting the existence of the phenomena, in this case the conscious mental states, themselves.

The first Epiphenomenalist response is that we should only reject the existence of entities that play no explanatory role, but that the mental does play some explanatory role. Mental and special science properties may play a non-causal explanatory role by capturing an important second-order or some other similarity among disjunct physical entities. Thus, Ockham's razor may not be sufficient reason to reject type Epiphenomenalism.

Ockham's razor may still be sufficient reason to reject Token Epiphenomenalism. On Token Epiphenomenalism, mental events cannot be causes. Furthermore, since mental events are tokens and cannot be shared, they cannot play the explanatory role of capturing similarities as mental types can. Thus, since mental events can play no explanatory role, we should not believe in them. This argument is sufficient for rejecting Token Epiphenomenalism about non-conscious events on a coarse-grained conception of events.

However, the Epiphenomenalist may argue that a fine-grained conception of events allows for mental events to be explanatorily relevant. If events are individuated by their properties, they may derive some explanatory relevance from their individuating properties. Whatever this explanatory relevance might be it clearly cannot be the same sort of relevance a causally relevant event would have despite the surface similarity of such explanations.

This explanatory role cannot help those who appeal to reasons explanations, however, since reasons are particular justifications for particular actions. I conclude that Davidson is wholly successful in his criticism of those philosophers advocating a non-causal role for reasons-explanations. Nonetheless, we still may have reason to believe in the mental insofar as it captures similarities among physical states. Mental properties may be relevant for this reason. Mental events may be individuated by these properties, and so they might be relevant as well. I will discuss knowing our own mental states as second-order properties two sections of this chapter hence to see whether this response is effective.

The second Epiphenomenalist response is that conscious mental states differ from the properties or events adverted to in the special sciences. Consciousness, arguably, does not play any role in scientific explanations, so an appeal to simplicity of explanation is no reason to conclude anything about its causal relevance or existence. Moreover, we do not know about our own consciousness because of the explanatory role it may play but because it self-evidently exists as our mental life itself. Our conscious states, thus, provide reason not to reject the existence of epiphenomenal mental entities. Consciousness may also, however, provide a different reason to reject Epiphenomenalism.

Our direct knowledge of our own minds suggests two further arguments against Epiphenomenalism. The first of these arguments, argument 6 in the list above, is that

Epiphenomenalism is incoherent⁶. The second of these arguments, argument 7 above, is that that we could not know or refer to our minds if Epiphenomenalism is true.

The argument that Epiphenomenalism is incoherent proceeds as follows. If Epiphenomenalism is true, then one cannot know that it is true or truthfully state it. If Epiphenomenalism cannot be known or truthfully stated, then we have no reason to accept it. Thus, we have no reason to accept Epiphenomenalism. Although this argument is not convincing, it will raise considerations about our knowledge of and reference to our own minds that will ultimately merit rejection of Epiphenomenalism.

4.7 Arguments against Coherence of Epiphenomenalism

The argument for the incoherence of Epiphenomenalism has at least two distinct forms. One argument is that Epiphenomenalism is, one might say, epistemically incoherent. This argument is put in terms of knowledge of our own minds or formation of beliefs about our own minds. The second argument is that Epiphenomenalism is referentially incoherent, and is put in terms of reference to our own minds. Neither of these arguments prove that Epiphenomenalism is incoherent.

The epistemic argument proceeds roughly as follows. If Epiphenomenalism is true, then one cannot know that one's mind is causally inert. If one cannot know that one's mind is causally inert, then one cannot know that Epiphenomenalism is true. Therefore, if Epiphenomenalism is true, then one cannot know that it is true. Finally, if one cannot know that something is true, then one should not believe that it is true. Therefore, one should not believe that Epiphenomenalism is true.

⁶ McLaughlin (1994) presents an argument against the coherence of Epiphenomenalism. He argues that Epiphenomenalism cannot be truthfully stated because the act of making a statement requires that one's belief causes one's utterance, so, if Epiphenomenalism is true, it cannot be stated. McLaughlin notes that this argument is ineffective because the Epiphenomenalist could easily doubt the causal theory of

The problem here is not a question of whether one would be justified in believing that Epiphenomenalism is true. It may be that we should believe even if we cannot know it. Perhaps if a theory is incompletely justified, we should believe that theory even if we cannot know it. The problem this argument presents is rather that one could not even come to have the belief that Epiphenomenalism is true if Epiphenomenalism is true.

The second argument for incoherence of Epiphenomenalism is an argument for referential incoherence. If Epiphenomenalism is true, then we cannot refer to our own minds. If we cannot refer to our own minds, then when we attempt to state the thesis of Epiphenomenalism, then we refer to nothing when we state it. When we say that minds are causally irrelevant, the term, "minds," does not refer to anything. And if a statement includes a term that fails to refer, the statement is meaningless. Thus, Epiphenomenalism cannot be stated meaningfully, and any theory that cannot even be stated meaningfully cannot be true. Thus, the Argument from Referential Incoherence has a stronger conclusion than the epistemic version, that Epiphenomenalism cannot be true.

The Epiphenomenalist can raise several objections to these arguments. First, the Epiphenomenalist might claim that the arguments incorrectly assume simple causal theories of knowledge and reference. Admittedly, if Epiphenomenalism is true, and knowledge, belief formation or reference requires that some entity cause our belief or the putative act of reference in the right way, then we cannot know or refer to our own minds. In the next two chapters, without assuming a *simple* causal theory of knowledge, belief formation and reference, I will argue that Epiphenomenalism is inconsistent with knowledge of and reference to our own minds. But before raising that issue I will argue that the above

statements (and action). Thus, I will not pursue this argument here.

arguments for incoherence of Epiphenomenalism each involve a scope ambiguity that renders them unsound.

Each of the above arguments involves a scope ambiguity. Each argument requires a premise that, on one reading, may be true but is harmless to the Epiphenomenalist, but under the required reading, is false. Thus, the arguments are unsound either because they are invalid, when assuming the (possibly) true reading, or because they have a false premise.

In the Argument for Epistemic Incoherence, the second premise involves a scope ambiguity. The relevant premise is that if Epiphenomenalism is true, then one cannot know one's own mind. The claim, 'We cannot know that our minds are causally inert,' can be understood in two ways. First, it could mean that we cannot know of our minds, of those very things, that they are causally inert. Second, it could mean that we cannot know the proposition that our minds are causally inert. It is true that we cannot know, assuming a causal theory of knowledge, of some particular causally inert entity that it is inert. However, that claim is not necessary for the Epiphenomenalist to be said to know his own theory. All the Epiphenomenalist needs to know is the proposition that minds are causally irrelevant. The second reading of the controversial claim is false. If Epiphenomenalism is true, we can know that proposition.

The Argument from Referential Incoherence involves a similar scope ambiguity in its premises. The relevant premise of this argument is that if Epiphenomenalism is true, we cannot refer to our own minds. This premise also involves a scope ambiguity. The claim could be interpreted in one of two ways. We could take it as meaning that of those things, our minds, we cannot refer to them. Or it could mean that we cannot state the following proposition: we cannot refer to our own minds. The first reading is true if

Epiphenomenalism and a simple causal theory of reference are true. If Epiphenomenalism is true, then, of those things, we can say nothing because we are not related to them.

However, this reading is not the one we need to state Epiphenomenalism. To state Epiphenomenalism, we need only state a proposition, the proposition that minds are causally irrelevant. If we say that need only take the second reading of the controversial premise. If Epiphenomenalism is true, we cannot state the proposition: minds are causally irrelevant. This reading is the only one the Epiphenomenalist needs to state her case, but this reading is clearly false. One can state a proposition in which one of the terms is causally inert.

Assuming a simple causal theory of knowledge, one cannot know of a particular, either a concrete or abstract particular (an entity or one of its particular properties, like its particular *redness*), with which one has no acquaintance that it is causally inert. Since, according to Epiphenomenalism, one's mind can never cause one's beliefs, on this view, one cannot know of one's mind that it is irrelevant.

Similarly, following a simple causal theory, one cannot say of a particular (rigidly designated) unicorn that it is causally inert or even that it does not exist. Of that particular unicorn one can know nothing because that unicorn does not exist, and one can know nothing of or refer to a nonexistent particular. To take another example, one cannot know of phlogiston, of that natural kind, that it does not cause combustion. No phlogiston exists for anyone to know these facts about. So, on the narrow scope reading of these claims, *prima facie*, we cannot know or refer to those entities⁷.

⁷ These examples may instead be taken to show that causal theories of knowledge and reference are false. I do not wish to endorse this argument here although it does nothing to endanger my argument. If these causal theories are false, then the incoherence arguments fail anyway. However, even if we accept these causal theories, the incoherence arguments will fail.

I do not want to argue that these claims are true. For the moment, I am willing to concede these claims, but only for the purpose of arguing that even if these claims are true, the argument does not go through. In the next two chapters I will argue that even on a less-simple causal/nomological theory of reference, we cannot be acquainted with our own minds, and for this reason Epiphenomenalism cannot be true. The reasons I have given in the preceding paragraph are too simplistic to establish such a conclusion. However, I intend to accept them for the sake of argument for the rest of this chapter.

A different reading of the relevant claim is all that is necessary to know, believe or claim that Epiphenomenalism is true. One can know the proposition that minds are causally inert just as one can know the proposition that unicorns do not exist, that unicorns have one horn, or that phlogiston is not the cause of combustion. To belabor the point, we can meaningfully state, and know, truths of number theory even though we are not causally related to numbers or sets. None of these claims requires acquaintance with any particular for one to know that it is true. These claims only require that one know a proposition that minds are causally inert, that unicorns and phlogiston do not exist, or that elementary number theory is true.

I will state the two arguments more formally to show how the scope ambiguity arises and how that ambiguity renders the arguments unsound. The first form of the argument is as follows.

- (1) If Epiphenomenalism is true, then for any person S and any mind x, such that x is S's mind, it is not possible that S knows (or comes to have beliefs about) x.
- (2) For any person S and any mind x, such that x is S's mind, if it is not possible for S to know (or come to have beliefs about) S's mind, then it is not possible that S knows (or comes to have beliefs about) that x is causally inert.
- (3) Therefore, if Epiphenomenalism is true, then, for any person S, it is not possible that S knows (or comes to believe) Epiphenomenalism.

- (4) For any person S, if it is not possible that S knows (or comes to believe) a theory, then we should reject that theory.
- (5) Therefore, if Epiphenomenalism is true, we should reject Epiphenomenalism.
- (6) If Epiphenomenalism is not true, then we should reject Epiphenomenalism.
- (7) Either Epiphenomenalism is true or it is not true.
- (8) Therefore, we should reject Epiphenomenalism.

This first argument for incoherence of Epiphenomenalism purports to show only that we should not believe Epiphenomenalism because even if it were true, we could not know or come to believe it. And, given our goal of acquiring truths, we should reject any theory that is false. Thus, the argument from (5) to (8) is uncontroversial and valid.

However, the inference from (1) and (2) to (3) is invalid with this premise (2). This version of premise (2), which we can call the wide-scope reading because the scope of the quantifier x ranges over the whole conditional, is arguably true on a generally causal account of knowledge. However, (3) does not follow from (1) and (2). (3) states that S cannot know that Epiphenomenalism is true. However, all that follows from (1) and (2) is that, of x, S's mind, S cannot know that x is causally irrelevant. But that claim is not equivalent to the claim that Epiphenomenalism is true. Given the examples of unicorns, phlogiston, number theory, one can know a proposition or set of propositions without being acquainted with any particular that fits into those propositions. The theory of Epiphenomenalism is, presumably, a proposition or set of propositions, and S can easily believe or perhaps even know that that theory is true without being acquainted with his own mind.

The argument for Epistemic Incoherence requires a different version of premise (2) to be valid. The argument for (3) can be given as follows.

- (1) If Epiphenomenalism is true, then for any person S and any mind x, such that x is S's mind, it is not possible that S knows (or comes to have beliefs about) x.
- (2') For any person S, if it is not possible for S to know (or come to have beliefs about) S's mind, then it is not possible that S knows (or comes to believe) the

proposition that, for any person T and any x such that x is T's mind, x is causally inert.

- (3) Therefore, if Epiphenomenalism is true, then, for any person S, it is not possible that S knows (or comes to believe) Epiphenomenalism.

The revised premise (2') makes the argument valid. Assuming that the claim *for any person T and any x such that x is T's mind, x is causally inert* is equivalent to the thesis of Epiphenomenalism, then (3) follows from (2'). However, this revised argument is unsound because premise (2') is false. We might call (2') the narrow scope version of the premise because the scope of the quantifier, x, covers only the predicate, *is causally inert*. As noted, one can know a proposition about particulars, with which one is not acquainted. I can know the proposition *Unicorns do not exist* without being acquainted with any particular unicorn. In fact, if I were acquainted with such a unicorn, then I couldn't know that proposition. So, the Argument for Epistemic Incoherence is unsound either because it is invalid, under the wide scope premise, or has a false premise under the narrow scope reading.

The second argument for incoherence of Epiphenomenalism purports to show that Epiphenomenalism is referentially incoherent. This argument involves the same type of scope ambiguity. It proceeds as follows.

- (9) If Epiphenomenalism is true, then for any person S and any x such that x is S's mind, it is not possible for S to refer to x.
(10) For any person S and any x such that x is S's mind, if it is not possible for S to refer to x, then it is not possible that S meaningfully assert that x is causally irrelevant⁸.

⁸ There is an additional response the Epiphenomenalist might make to this premise. She might claim that, following the causal/historical theory of reference apparently assumed, when S makes claims about minds, S refers to something other than minds, say brains. Thus, when S claims that minds are causally irrelevant, his term "minds" actually refers to brains, and so S has said something meaningful albeit false.

This response has two problems. First, it seems false that S refers to brains when he uses the term "mind." We would not wish to say that chemists who used the term "phlogiston" were in fact referring to oxygen when they used that term. Such an interpretation would seem to make them correct when they were in fact wrong to say, for example, that phlogiston causes combustion. Cases like these are something of a puzzle for the causal/historical account because the referent is not the cause. A causal/historical theorist would respond that being a cause is a necessary but not a sufficient condition for reference. Questions about the simple application of a causal/historical account will be discussed in the next two chapters.

- (11) Therefore, if Epiphenomenalism is true, then it is not possible for any person S to meaningfully state it⁹.
- (12) Any theory that cannot be meaningfully stated cannot be true.
- (13) Therefore, Epiphenomenalism cannot be true.

This Argument for Referential Incoherence is unsound just as the Argument for Epistemic Incoherence was. In this argument, we may call premise (10) the wide scope claim because the scope of the quantifier x ranges over the whole conditional. This wide scope reading is apparently true on a causal/historical account of reference. One cannot, on this view, refer to some particular with which one cannot be causally related. But (12) does not follow from premises (9) and (10). The reason is that we can make meaningful statements without referring to any particular. For example, we can meaningfully make claims about unicorns and phlogiston without being causally related to them. So, the argument for (11) that assumes (10) is invalid.

The Argument for Referential Incoherence of Epiphenomenalism would be valid but unsound with (10'), a narrow scope reading of (10).

- (9) If Epiphenomenalism is true, then for any person S and any x such that x is S's mind, it is not possible for S to refer to x.
- (10') If, for any person S, it is not possible for S to refer to x, such that x is S's mind, then it is not possible that S meaningfully assert the proposition that x is causally irrelevant.
- (11) Therefore, if Epiphenomenalism is true, then it is not possible for any person S to meaningfully state it.

The second problem with this response is that it does not appear to help the Epiphenomenalist. On this view Epiphenomenalism could not be stated truthfully because the basic statement of Epiphenomenalism, that minds are causally irrelevant would become false. If Epiphenomenalism cannot be stated truthfully, then that still seems reason to reject it.

⁹ An additional point to note is that Epiphenomenalism is not a claim about only one's own mind, but is a claim about anyone's mind. Since the epistemic and referential considerations apply equally well to anyone's mind, it should be easily possible to state these arguments in more general terms. I state the arguments only in terms of first-person knowledge and reference because our ability to know and refer to our own minds is, I hope, uncontroversial, and because I will focus on the first premise of each of these arguments in the next chapter.

However, premise (10') is false. As we noted above, we can make meaningful, even true, statements about particulars to which we cannot refer. So, on either scope reading of the appropriate premise, the Argument for Referential Incoherence is unsound.

Perhaps it will be argued that these examples I appeal to are relevantly different from the case of Epiphenomenalism since they all involve nonexistent or fictional entities. Traditionally philosophers have had difficulty dealing with nonexistent entities, but our ability to make coherent and true statements about them is uncontroversial if puzzling. My argument appeals to nothing more controversial than the fact that we can speak and know about non-existent objects. Thus, epiphenomena and nonexistent entities should not be disanalogous in this respect. Furthermore, it is possible to give examples of abstract entities about which we can know or make meaningful statements. I did not emphasize an analogy with abstract objects because they represent a problematic class of apparent counterexamples to causal theories of knowledge and reference. We are able to know propositions or make meaningful statements about fictional or nonexistent entities, and Epiphenomenalism need appeal to nothing more than a similar ability for the case of causally irrelevant particulars.

This reasoning shows that there is no difficulty with coming to believe that Epiphenomenalism is true. However, knowing is more than having a true belief. Ignoring Gettier problems, we may be safe in assuming that justification is also required for knowledge. In this case, the Epiphenomenalist may be justified in holding his belief by appeal to the Exclusion Argument itself. The Exclusion Argument certainly appears to be sound, and believing something on the basis of a sound argument seems as good a justification as one might need. Therefore, if Epiphenomenalism is true, there is no reason to think Epiphenomenalism incoherent. We should presume in favor of coherence of a theory

unless it can be shown not to be coherent. Since these arguments have failed to show Epiphenomenalism incoherent, we should presume that it is coherent.

Since, as I have argued in this chapter, none of these brief arguments against Epiphenomenalism is effective, we must appeal to a more careful argument to show that Epiphenomenalism is false. In the next two chapters, I will argue that problems with knowing and referring to one's own mind effectively refute Epiphenomenalism. In effect, I will argue against Epiphenomenalism based on a causal or nomic theories of knowledge and reference for a sophisticated version of the controversial premises assumed in the arguments for Epistemic Incoherence and Referential Incoherence of Epiphenomenalism. This premise will conflict with our uncontroversial ability to know and refer to our own minds.

CHAPTER 5

KNOWING NONCONSCIOUS MENTAL ITEMS

5.1 Introduction

In the last chapter I claimed that many arguments against Epiphenomenalism were not sound. In this chapter I will discuss problems Epiphenomenalism might have with first-person knowledge and reference. I will argue that Epiphenomenalism can avoid many of the problems that have been thought to defeat it by appealing to a metaphysical, nomic or causal connection between the mental and the physical. I will conclude this chapter by arguing that the relation, whatever its variety, of the mental to the act of reference or the our beliefs is not one that grounds reference, belief formation or knowledge.

Although Epiphenomenalism is coherent, it does raise problems for knowing or referring to our own minds¹. Let us return to the question of whether we could know our own minds if Epiphenomenalism is true. I assumed for the sake of argument, in argument 6 in the previous chapter, the Argument against the Coherence of Epiphenomenalism, that we could not know our own minds if Epiphenomenalism were true. Although the Argument against the Coherence of Epiphenomenalism is not sound, problems of knowledge and reference to our own minds make it unlikely that Epiphenomenalism is true.

Two varieties of mental entities (identified by our epistemic relation to them) are relevant to our discussion here. The first are mental entities that can be known only inferentially or indirectly. Propositional attitudes, cognitive processes and unconscious beliefs and desires are arguably of this type. I will call these mental entities NCEs for non-

¹ Frank Jackson (1982), for example, thinks that problems with coming to know one's mind show that the mental must be relevant for the production of other mental properties or states. Jackson thinks that the mental is not completely causally inert, only unable to have any physical effect. Thus, although he claims to defend Epiphenomenalism, he defends only a restricted version of Epiphenomenalism, not the robust

conscious entities. The second type are mental entities that are at least sometimes known directly. Conscious mental states are of this type. I will call these *qualia*, following the common philosophical terminology for conscious states or the essential character of those conscious states.

5.2 Existence of Nonconscious Mental Entities.

I will first pursue how Epiphenomenalism might deal with problems of knowledge of and reference to NCEs. I will argue for the causal relevance of the mental from our ability to know or come to have beliefs about our own minds. I assume that some NCEs exist and that we do know or come to have beliefs about at least some of these mental entities. First, I will argue that Token Epiphenomenalism is not consistent with our knowing or having beliefs about NCEs, and so, Token Epiphenomenalism is false. Second, I will discuss Block's (1990) limited form of Epiphenomenalism about second-order properties and argue that it is incompatible with the exclusion argument. Third, I will argue that Block's notion of a second-order property as characterizing mental properties allows Type-E to be compatible with us to knowing our own NCEs inferentially.

The terms 'NCE' and 'qualia' in the following arguments are intended neutrally assuming nothing about the kind of Epiphenomenalism under discussion. When it becomes relevant, I will apply these criticisms of Epiphenomenalism to type Epiphenomenalism. In addition, these terms are not meant to imply anything about the possibility of a mental item's independent existence. These items may be independently existing objects or substances, or they may be different properties of the same object that depend for their existence on that object. In short, nothing significant yet depends on the choice of terminology; convenience

Epiphenomenalism here at issue.

and generality are the only reasons for this terminology. When it is important to distinguish the varieties of entities and Epiphenomenalisms, I will do so clearly.

One argument against Epiphenomenalism is based on our knowledge of our own minds. This argument proceeds as follows.

- (1) If Epiphenomenalism is true, it is not possible for us to know our own minds or to form beliefs about them.
- (2) We do know our own minds or to form beliefs about them.
- (3) $\therefore$ Epiphenomenalism must be false.

Statement (1) is uncontroversial, but statement (2) requires considerable argument. There seem to be several possible ways for us to know our NCEs. They may fill some inferential or explanatory role for them to ground our beliefs; they may be related to us in the right causal or nomological way; they may be known as we know nonexistent or fictional entities. According to the first premise, if Epiphenomenalism is true, it is impossible to know our NCEs in any of these ways.

That we know our mental processes or properties by the inferential or explanatory role they play contradicts standard philosophical wisdom that our knowledge of our own minds is direct and non-inferential. However, the view that our knowledge of at least some kinds of mental states is inferential is fairly well-established.

It is fairly commonplace to note that the heavy lifting of many mental processes are unconscious². For example, language processing, perceptual processing, including face recognition, and other perceptual and motor processes occur entirely unconsciously. We are conscious only of the end product of these mental processes.

Nisbett and Wilson (1977) argue persuasively that our knowledge of at least some kinds of mental entities, in particular (what we would ordinarily consider) causes of our

² This claim is often referred to as "Lashley's Observation," from Lashley (1956)

actions and decisions, and the cognitive processes (putatively) underlying our behavior, is based on inference from our actions and background (putatively) causal theories of action. For example, Nisbett and Wilson found a handedness bias in reported preference for identical nylon stockings. Subjects rated the best of a set of stockings, all of which, unknown to the subjects, were identical. The subjects gave the highest rating significantly more often to the pair that corresponded to the subjects' dominant hand. Right-handed subjects, for example, were more likely to say that the pair placed on the right was the best pair of stockings.

When confronted with this fact about their decision-making, subjects generally responded with incredulity. So, although there is a lively debate in the psychology literature about the extent of our access to these processes (and to what extent we merely forget the determinants of our decision-making), it seems undeniable that at least some aspects of the (putative) causes of our behavior are not directly available to or introspectible by us.

In addition, propositional attitudes are at least sometimes known only indirectly even in the first person case, except in cases of occurrent thoughts. First-person attribution of propositional attitudes is generally taken to be authoritative, but, without some associated phenomenology, it is a mystery how such knowledge could be direct. One explanation for how we make such attributions is that we make them by inference. In addition, in learning folk psychology, for example, children appear to make inferences in attributing propositional attitudes and intentional states to themselves (Gopnik 1996)³. These claims are not entirely uncontentious, and I cannot argue more for them here, but they do make a strong case that our access to at least some of our mental entities is inferential. Arguing that these items

³ These attributions may become automatic over time so that we may not think of them as inferences, but they still do not appear to involve the kind of direct acquaintance that we have with our conscious mental

cannot be known if Epiphenomenalism is true constitutes a meaningful critique of Epiphenomenalism, at least for these kinds of mental states. Nonetheless, for those philosophers not convinced by the preceding evidence, I will later argue that conscious states cannot be known if Epiphenomenalism is true.

We have two ways of knowing our own NCEs by inference. The first possibility is that we know our NCEs in the same way we know unobservable entities postulated by a scientific theory. The second possibility is that we know our NCEs, in the same way we know abstract objects, by their picking out similarities in the world.

The first possible way of knowing our own mental entities is suggested by Nisbett and Wilson. They claim that first-person knowledge of cognitive processes is a variety of inference to the best explanation. At least some mental entities are, on their view, unobservable entities required for our theory to explain our behavior.

However, Epiphenomenalism gives no grounds for believing in these mental entities because of their explanatory role in an inference to the best explanation. A realist about unobservable or theoretical entities believes in them because they unify the observable phenomena.⁴ An unobservable entity that unifies phenomena is one which we believe is a single cause for a diverse range of experimental evidence. For example, the electron is an unobservable entity whose existence we infer because it unifies a range of experimental evidence from the movements of charged oil droplets in Millikan's experiment to the conduction of electric current to the occurrence of lightning. The electron is the single cause of these diverse phenomena. However, since epiphenomena cannot be the cause of any

states.

⁴ An anti-realist about unobservables would, of course, find the mental entity's explanatory role no reason to believe in its existence. So, for this case, we need only worry about the realist about unobservables. In any event, I assume for the sake of this work that scientific realism is true.

phenomena at all, they cannot be the unifying cause in the explanation of our own behavior. Therefore, one cannot know one's own NCEs in the way that we know unobservable or theoretical entities if Epiphenomenalism is true.

The Epiphenomenalist may respond that although epiphenomena could not fill the role of an unobservable entity, epiphenomena could be effects of some causally relevant entity, say a physical or neural entity, that did unify the observable phenomena. The mental entity may, thus, be known as an effect of the theoretical entity that does play a role in explaining the phenomena. This response may describe a genuine possibility, but it gives no reason to believe in the NCE. Without already knowing that there is a relation between the NCE and the physical entity, belief in the physical entity would give no reason to believe in the NCE. Therefore, on this view, we are trapped in a vicious circle; we cannot know our own NCEs without already knowing our own NCEs. So, the possibility that our mental entities are effects of an unobservable entity will not provide reason to form beliefs about our own mental entities.

The second possible way of knowing our own minds, that we know mental entities in the same way we know abstract entities, offers more hope. Although problems with whether and how we can know abstract entities are the primary reason many philosophers question the existence of abstract entities, if we have the same grounding for our beliefs in NCEs that we have for our beliefs in abstract entities, then knowing our NCEs does not present a special problem for Epiphenomenalism. Abstract entities in general play an important explanatory, but non-causal, role that may ground our belief in them by explaining similarities among particular objects. Red objects are red, one might think, because they share the universal property of *redness*. Thus, the view that NCEs can be known because of

the similarities they explain can avoid the problem of knowing our NCEs by incorporating it into the general problem of knowing abstract entities. If we can give the same reasons for believing in NCEs that we can give for believing in abstract objects, then Epiphenomenalism does not have any special problem.

At this point Broad's (and McLaughlin's) distinction between Type-Epiphenomenalism and Token-Epiphenomenalism becomes important. Token-E cannot allow for the possibility that we know NCEs because of similarities they explain. On Token-E, the epiphenomena are events or singular occurrences, not properties. Singular occurrences clearly cannot be adverted to in an inference based on a shared similarity. If, as I have argued, the only way to know our own cognitive processes and propositional attitudes is by appeal to similarities among different types of beings with minds, then Token-E is incompatible with our knowing at least some of our own mental events. And, as I am assuming, since we can know or come to have beliefs about these mental events, Token Epiphenomenalism must be incorrect, for the case of inferentially known entities. In the next chapter I will raise the issue of knowledge of conscious states for both Token-E and Type-E. But, assuming we can know all mental tokens, the case of mental tokens that we know only inferentially shows that Token-E cannot be true for these mental tokens.

Type-E can, on the other hand, appeal to the possibility of shared similarities among beings with minds since, according to Type-E, the epiphenomena are properties. A purely physical description of a human brain might be enough to explain human behavior, but it would not be enough to explain similarities to the behavior of other beings with mental lives. A characterization in terms of functional roles, for example, may be necessary to explain the similarities we can see or which seem possible among the behavior of beings with

heterogeneous physical realizations, e. g. octopus brains or silicon based Martians. If Type-E is true, appeal to NCEs can still play an explanatory role by picking out some similarity among beings that all share a mental property.

At first glance, we might think that appealing to a functionalist account of mental similarities conflicts with Epiphenomenalism. If Epiphenomenalism is true, mental entities do not have causal powers. If minds do not have causal powers, the only similarity of causal powers they have is that they have none. Therefore, if the similarity among minds from one species or individual to another is the causal role of the entity, then Epiphenomenalism does not allow us to know our own NCEs by an inference to the similar causal powers they have in different species or individuals.

5.3 Block's Theory

Block (1990) argues, to the contrary, that mental terms categorize causally relevant physical properties without being themselves causally relevant. Block's view is that mental properties are second order properties of physical entities. A good explanation of second-order properties comes from Kim.

Let *D* be a set of ('first-order') properties: a *second-order property* over *D* is the property of having some property in *D* satisfying a certain specification *C*. Where *C* involves causal relations (that is, *C* specifies a 'causal role'), we may call the second-order property a *functional property*. Properties in *D* satisfying *C* are the *realizers* of the second-order property in question. (1997, 290)

Functionalism in the philosophy of mind appears committed to the view that the mental properties are a variety of second order properties. If the only way to capture the similarities among beings with mental properties is by taking mental properties, at least cognitive processes and propositional attitudes, to be functionally defined, then these mental properties must be a variety of second-order property.

Block (1990) argues for the skeptical claim that functional properties in general are not “causally relevant to the effects in terms of which they are applied.” (Block 1990, 156) Block illustrates this claim with the example of the second-order property of dormitivity. Dormitivity is “the possession of some property or other (for example, a first-order chemical property) that is causally relevant to sleep. That is, x is dormitive = x has some property that is causally relevant to sleep (when x is ingested).” (Block 1990, 155) However, dormitivity is not itself causally relevant to making one sleep; in general the first-order chemical property alone is causally relevant to making one sleep. If Block’s claim is correct, then mental properties are in general causally irrelevant to the effect in terms of which they are defined, but mental properties can still be explanatorily relevant.

Block’s suggestion⁵ is a limited form of Epiphenomenalism. He does not commit to the view that second-order properties are completely irrelevant but only generally causally irrelevant to the effects that are adverted to in defining the property. The second-order property can be causally relevant to the knowledge that the pill puts one to sleep. The causal inertness of the mental only extends to the effects the second-order property is defined in terms of, so the mental can still be relevant for our knowing our own mental states. Thus, Block’s theory might provide a way for the mind to be known without thinking that it is causally relevant to first order properties.

Block’s restricted Epiphenomenalism does not, however, solve the problem presented by the Exclusion Argument. According to the assumptions of the Exclusion Argument, some physical state or event determines our knowledge of the second-order property. Some physical property, perhaps the writing on the box of the pharmaceutical,

⁵ Block seems to take this variety of Epiphenomenalism as a skeptical challenge to be overcome rather than a position he wishes to defend.

produces a brain property that determines our knowledge. And, according to the Exclusion Principle, since some physical property determines our knowledge, no other property can cause the knowledge. If this conclusion is correct, then it must be that in general we do not know our mental properties by their causal relations to us. Without rejecting the Exclusion Principle, one cannot accept Block's limited form of Epiphenomenalism. Therefore, for Type-E to be consistent, it must be that our mental properties are completely causally inert, and if we are to know our mental properties at all, it must be in some non-causal, or indirectly causal, way.

The argument is that Block's position will not provide a solution to the problem of mental causation as it is raised by the Exclusion Argument (although Block is not concerned with the Exclusion Argument). To accept Type-E as a solution to the problem of mental causation, we must accept that second-order properties cannot be causally relevant to anything, including our knowledge of them. Nonetheless, the second-order property conception of the mind provides a way of knowing our own minds.

5.4 Knowledge of the Mental as Second-Order Properties

The notion of mental properties as epiphenomenal second-order properties makes knowledge of one's own mind seem mysterious. Appeal to the property of *dormitivity* does not appear to explain much on its own since it explains why some substance puts one to sleep by appealing to its dormitive virtue, its capacity to put one to sleep. Nonetheless, appealing to second-order properties can avoid vacuity if the appeal is to a complex web of propositions, an autonomous science, that ultimately does not rely on any underlying physical structure. Various drugs that act by distinct chemical means but which all have the effect of putting people to sleep really do have something in common. A science of

psychology that applied both to putative Martians with silicon-based brains and to us would, presumably, involve appeal to causally individuated second-order properties.

The problem, then, is to explain how, if Epiphenomenalism is true, we can know this similarity among first-order properties. In an ordinary case of seeing a similarity among objects with a certain property, say *redness*, we see the token of the property, the redness of *that* apple, and notice that *that* redness is like another redness, say the redness of *that* tomato. Thus, ordinarily, to know a universal or a type, we need to know the property tokens first, then infer the existence of the type. However, in the case of Type-E, we cannot know the type by simple inference from the tokens. To know the second-order type, we must first have the concept of the second-order type, and then see that the particular first-order property meets the causal specification of the second-order type.

This theory of second-order properties raises two important questions. Can these epiphenomenal second-order properties be real? And is this theory compatible with Epiphenomenalism?

In answer to the first question, we should think that these properties are real insofar as we think the science that appeals to them is true. If we can specify the property without being acquainted with its tokens, then the property can be causally inert. And it does seem possible to specify properties without being acquainted with any instances of them. For example, I can, more or less arbitrarily, specify the property of having the shape of a chiliagon, to borrow Descartes' example, without having any acquaintance with chiliagons. I can specify the property of being the third person mentioned in the Bible without having any idea who that person might be. These examples show that we can arbitrarily invent properties (or perhaps less contentiously second-order predicates). Some of the properties

specified in this way may not be real, but one way, but not the only way, for those properties to be real is if they fit into a true scientific theory. This answer to the question may seem question-begging--the theory can only be true if the properties adverted to in it are real. However, we have ways of testing for the truth of scientific theories, for example, by empirical hypothesis testing, that do not require us first to know the reality of the properties. Thus, if functionalism (or any second-order property account) can provide true explanations⁶, then the second-order properties themselves can be real.

The second question about second-order properties is whether their reality is consistent with Epiphenomenalism. These properties, even properties that are individuated by their causal roles, can be epiphenomenal. Second-order properties, I have suggested, are known by one's first having a concept of the property, and then matching the causal properties of the event with the second-order concept. So, second-order properties need not have causal power themselves. If we can specify the concept independently of any experience with instances of the property, then there is no reason to think the second-order property must be causally relevant.

Thus, the notion of second-order properties provides a way that we might know the mental if the mental is epiphenomenal. The conception of the mental as epiphenomenal is, therefore, consistent with our knowing our own minds (at least some mental properties) by inference. So, this conception of the mind counts against argument 4, the Argument from Simplicity, and argument 5, Davidson's Argument. Davidson's Argument was supposed to show that the mental had to be causally relevant for it to play any explanatory role. And the

⁶ A possibility very much in doubt. Block (1978) argues that functionalism cannot find a middle-ground that is neither too chauvinistic, arbitrarily limiting minds to humans, nor too liberal, allowing things that clearly have no minds to be judged as having minds. So, functional characterization of the mental may simply fail as a philosophical theory of the mind.

Argument from Simplicity was supposed to show that, because epiphenomena could play no explanatory role, we should reject the existence of the mental if Epiphenomenalism were true. Since the second-order property conception of the mind allows for the mental to play a non-causal explanatory role, the Epiphenomenalist can reject a crucial premise of the above argument. So, Davidson's Argument and the Argument from Simplicity do not show that we should not believe in the existence of the mental if Epiphenomenalism is true.

Reasons or other mental explanations may fulfill some explanatory role for particular events if reasons are not themselves causally relevant. The explanatory role played by a second-order property can capture some similarity among events, and by fitting those events into a larger scheme, the second-order property can explain why a particular event's occurrence counts as an event of that type. For example, Epiphenomenalism can provide an explanation of why Fred's action of getting a beer counts as an action of beer-getting. Fred's brain event that realizes his desire to get the beer can be seen to have a second-order property, some non-physical similarity with other physical events. These other physical events resulted in acts of beer-getting *ceteris paribus*, but such an explanation need not explicitly advert to the causal powers of the realizing brain-state. So the second-order property view can explain why Fred's behavior is one of getting beer by an appeal to a general theory of beer-getting that makes no assumptions about the physical realization of Fred's beer-desires. So, the second-order property theory of the mental allows for genuine explanations that do not require causal relevance of the second-order properties themselves.

The conception of the mental as second-order properties provides a way we could know, by inference, some of our own mental properties if Epiphenomenalism is true. If the

mental properties fit into explanations like that of Fred's beer-getting behavior, Fred can know his own mental state, his beer-desire, by appealing to that theory of beer-getting behavior. He can see his own action of getting beer, and infer that this action was caused by a brain state, which he infers is an instantiation of a beer-getting desire, and so he knows or has reason to believe he was in a beer-desiring state. This explanation is too trivial to be truly convincing, but if Fred's beer-getting desires fit into a complex web of common-sense or scientific psychology statements, then the explanation can be less than trivial. For example, Fred has a theory of mind that desire-states are instantiated by states that, *ceteris paribus*, cause attempts to get the object desired. Fred's theory, then, would include some specifications of the *ceteris paribus* conditions under which one would exhibit beer-getting behavior, and these conditions could be specified (although not completely) in terms of second-order properties like beliefs and other desires. So Fred's explanation, by fitting into a complex common-sense or scientific psychology can genuinely explain behavior without assuming causal relevance of the beer-desire itself. So, Epiphenomenalism does not raise insurmountable problems of inferential knowledge of our own minds.

This conception of how we know our minds does not fit well with the views of Epiphenomenalists. Huxley (1898), for example, clearly viewed the mind as caused by the neurophysiology of the brain, but if the mind is a second-order property of the brain, the mind is not caused by the neurophysiology of the brain.

One response to this point is that we need not respect the intuitions of philosophers who have thought minds are effects. It may be that finding any solution to the problem of mental causation is sufficient reason to accept the analysis. Nonetheless, it may be possible to capture something of the original intuition about the dependence of the mind on the brain

within the second-order property account. Since the second-order property is defined in terms of the first-order property, it follows that the second-order properties will supervene on the first-order properties. Anything with the same first-order property will have the same second-order property. Although the dependence of the mental on the physical will not be causal dependence, it is true that any event will fall under a mental event type in virtue of falling under a physical event type. This variety of dependence may satisfy the intuition that the mental is determined by the physical even though that determination turns out not to be causal. However, even if this formulation does not satisfy the intuition, the motivation for finding a solution to the problem of mental causation may be reason to accept whatever form of Epiphenomenalism is available.

5.5 Conclusion

I have argued for three main theses in this chapter. First, because Token-E does not allow us to know some mental entities that we can know, Token-E is false. Second, Block's suggestion of a limited form of Epiphenomenalism does not offer an Epiphenomenalism that is compatible with the Exclusion Argument. However, third, his notion of the mental as second-order properties can make Type-E consistent with our knowing our own minds, specifically our non-conscious mental properties, by inference.

This solution must remain somewhat tentative because it depends on a theory of mental properties as second-order properties. This second-order property account may not capture the nature of conscious mental states. In addition, the second-order property account does not seem to fit well with the possibility that mental content is wide. If mental content is individuated nonlocally, then content properties cannot be simply properties of neural properties. For these reasons, we can take the second-order property account as only a

tentative view of the mind, and less compelling as a solution to the problem of causal relevance of mental content, and so this solution to the problems raised by Epiphenomenalism must be equally tentative.

The above arguments against Epiphenomenalism apply to our knowledge of others' minds just as they do to our knowledge of our own minds. In fact, they apply to any set of second-order properties. The main difference between these kinds of knowledge is that we tend to be more certain that we know our own minds than that we know others' minds or know other second-order properties. Epiphenomenalism raises problems unique to knowing our own minds from our direct knowledge of conscious mental states. In the next chapter I will turn to knowledge of one's own conscious states.

CHAPTER 6

FIRST-PERSON KNOWLEDGE OF CONSCIOUS STATES

6.1 Introduction

Consciousness, to paraphrase Nagel (1974), makes Epiphenomenalism seem really impossible. Consciousness appears to resist explanation in terms of second-order properties. More importantly and less controversially, we do not know or refer to our conscious states inferentially. Epiphenomenalism is intuitively appealing for consciousness because it is hard to see how consciousness could be explained in causal terms, but consciousness raises more problems for Epiphenomenalism.

In the previous chapter I argued that the problems with knowledge of our own minds could be overcome assuming a controversial theory of the mental as second-order properties, properties of physical properties. This view provided a tentative solution to the problem of knowing epiphenomenal mental states that we could reasonably be thought to know by inference. However, it is not in general true that we know our conscious mental states by inference. Thus, consciousness resists the solution an Epiphenomenalist might give along these lines. Problems of knowledge of, and reference to, our own minds continue to plague Epiphenomenalism about consciousness. I will argue that the Epiphenomenalist must reply to these objections by appealing to a nomic or causal correlation between qualia and the physical. Finally, I will argue that we have no reason to think the nomic or causal correlation envisioned by the Epiphenomenalist will be of the correct kind to ground knowledge or reference.

For simplicity, I will call the view that our conscious mental states are epiphenomenal 'Qualia Epiphenomenalism.' 'Qualia' is a term of art in philosophy of mind,

referring to the subjective character of our experience, the qualitative feel of our consciousness, or the feeling of what it is like to be in a conscious state. 'Qualia' is most generally used to describe properties of experience although sometimes it is used to refer to the experience itself. I will take the term to have either sense, and will distinguish the two possible senses if the need arises.

Philosophers who think that it is impossible to explain the nature of qualia in causal or functional terms tend toward Qualia Epiphenomenalism. If consciousness defies analysis in terms of functional roles, they argue, then it plays no functional role, and so is causally irrelevant¹. Thus, Qualia Epiphenomenalism is a view held by many philosophers independently of the Exclusion Argument. As such a view, Qualia Epiphenomenalism may provide an independently justified solution to the problem of mental causation. For this reason, it is especially important to consider whether Qualia Epiphenomenalism can be correct.

One criticism of Qualia Epiphenomenalism is that it is unfalsifiable², that no possible evidence would count against it. If I suddenly ceased to have conscious mental states, I would not thereby cease to have all the same beliefs about my mental states that I do. If the lack of consciousness could not convince me that I did not have them, then nothing could. The conclusion, then, is that Qualia Epiphenomenalism is not really an empirical hypothesis, and so should be rejected.

This criticism is incorrect for two reasons. First, the required counterfactual claim is false. It is not true that if I did not have conscious states, then I would still believe that I did.

¹ This criticism appears in Shoemaker (1975) as an argument for functionalism although Block (1978) argues that simply because a conscious state cannot be analyzed in functional terms does not imply that the conscious event is causally irrelevant. Finally, Chalmers (1996) accepts an argument like the one above while embracing consciousness Epiphenomenalism.

If consciousness is nomically correlated with physical states, then the physical states could not occur without the conscious states. It certainly seems to be true that consciousness is nomically correlated with neural states of the brain¹. Thus, the counterfactual premise necessary for the argument is false.

Second, falsifiability is a consideration only for empirical theories or explanations. We know our qualia but not, in general, as an explanandum. Three different justifications support this claim. (1) Mental phenomena are themselves our experiences. It is impossible to deny the existence of our experience. (2) Qualia, although mysterious, must exist because we have direct access to them; we must have qualia even if they explain nothing because we are directly acquainted with them. We know our minds immediately, through direct acquaintance. (3) Qualia are the data that we must begin with, the essence of our mental lives. Conscious phenomena are the means by which we perceive the world, our representations. These responses to the qualia skeptic all appear to be true, but they may not be compatible with Epiphenomenalism.

That conscious states and physical states may be nomically correlated shows that the argument against Epiphenomenalism must be formulated not as a claim about falsifiability of Qualia Epiphenomenalism but as a problem about our knowledge and reference to our own minds. The second two responses, (2) and (3), make claims about knowledge of and reference to our own minds⁴.

² This criticism appears in Lycan (1987) in defense of functionalism and in Smullyan (1980) as a criticism of dualism.

³ This example assumes that brain states are physical states. I assume this claim only for the sake of clarity. Presumably, consciousness would also correlate with whatever basic physical states underlie the brain states.

⁴ Tye (1995) gives perhaps the most detailed argument against Qualia Epiphenomenalism involving problems of knowledge. He argues that Epiphenomenalism makes the problem of other minds insuperable. It would be impossible, the argument goes, to know whether someone else is conscious since someone could always have the same causal relations with the observer without being conscious. Jackson's (1982) response to a similar

6.2 Problems with (1): Having Qualia

The claim (1) above does not directly involve knowledge of our own minds but it deserves mention as a difficulty that Epiphenomenalism encounters. The claim is that our conscious states, like pain, simply are our experiences, and, we cannot possibly doubt the existence of our experiences. This claim may be evidence against the eliminativist about conscious states, but it does not help the Epiphenomenalist. Our having experiences is, rather, a problem for Epiphenomenalism since having a concrete existent or property appears to require that existent or property to have causal power. If our experiences are not causally relevant, then it appears impossible for us to have them.

The Epiphenomenalist may respond that having a property does not require a causal relation between the object and the property that inheres in it. Thus, having an itch, for example, may just be the instantiation of the property *itchiness*. A relation to a universal, the property *itchiness*, cannot require a causal relation between the person instantiating the universal and the universal itself. However, one's having experiences is not simply a matter of instantiating a property. By having an itch, I instantiate the property *itchiness*, but in addition, I have a relation to a spatio-temporal entity, *the itch*. My relation to the type *itchiness* is one of inherence, but my relation to that particular itch is something more. The most plausible candidate for this relation is a causal one.

The Epiphenomenalist could argue that mental entities could be had by the subject in a noncausal way, without the subject being causally related to those qualia. The

problem is that one could know that others were conscious by analogy with oneself. This argument is, of course, Mill's analogical argument for the existence of other minds. One knows in one's own case that one is conscious, and, since others are relevantly like oneself, others are conscious.

Tye replies that Jackson cannot argue in this way because he cannot know in his own case that he is conscious. As already noted, Jackson is not committed to the view that qualia are completely causally irrelevant, only irrelevant to physical events. Thus, Jackson does not have the problem that Tye suggests, but the Epiphenomenalism under consideration here may have this problem.

Epiphenomenalist would then need to explain the relation in some way that did not merely assert that some mysterious noncausal relation holds between a person and her conscious states. Without a plausible explanation of this relation, accepting Epiphenomenalism requires believing a mystery every bit as large as the mystery of mental causation that Epiphenomenalism is an attempt to avoid. Certainly the Exclusion Argument seems to entail Epiphenomenalism, but perhaps mental causation actually occurs mysteriously without any explanation. This mystery would be no greater than accepting the mystery of a non-causal relation to our mental states, and so we would have no reason to accept Epiphenomenalism as a response to the problem of mental causation as it appears in the Exclusion Argument.

A better response for the Epiphenomenalist is to concede that my relation to my conscious states is causal but deny that this fact shows that Epiphenomenalism false. We might be causally related to a physical state that either causes or nomologically determines the conscious state. If the conscious state has a physical cause or determining property, then one could have the experience by having this physical state. We could have the conscious state by means of its physical determiner, which relates to the qualia causally or nomically. Thus, the physical cause or correlate of the conscious state could be a means to have experiences even if we are not directly causally related to the conscious states.

It appears, then, that the problem for the Epiphenomenalist of how we can have our mental states can be handled provided that mental entities can be causally or nomologically determined by physical entities. This claim in defense of Epiphenomenalism will be the common response to other problems I will raise against Epiphenomenalism. Although the Epiphenomenalist can hold this view of having one's qualia, it remains to be seen whether Epiphenomenalism allows one to know or refer to one's qualia.

The (2) and (3) justifications for belief in qualia raise problems for how we know or refer to our own qualia.

6.3 Problems with (2): First-Person Knowledge of Qualia

Response (2) is that we know our own minds because we are directly acquainted with them. The problem with this claim is that it seems impossible for us to be directly acquainted with our own minds if Epiphenomenalism is true. Thus, if we really are directly acquainted with our own minds, Epiphenomenalism is false. Acquaintance is a causal relation, yet Epiphenomenalism does not allow for our qualia to cause our beliefs. This raises one basic problem with knowledge of our own qualia that I will address immediately, and two further problems based on our purported privileged access to our own qualia and on the coherence of skeptical hypotheses. But first, I will discuss the basic argument. In the chapter four I discussed the following argument (renumbered for this chapter).

- (4) If Epiphenomenalism is true, then for any person S and any mind x, such that x is S's mind, it is not possible that S knows (or comes to have beliefs about) x.
- (5) For any person S and any mind x, such that x is S's mind, if it is not possible for S to know (or come to have beliefs about) S's mind, then it is not possible that S knows (or comes to have beliefs about) that x is causally inert.
- (6) Therefore, if Epiphenomenalism is true, then, for any person S, it is not possible that S knows (or comes to believe) Epiphenomenalism.
- (7) For any person S, if it is not possible that S knows (or comes to believe) a theory, then we should reject that theory.
- (8) Therefore, if Epiphenomenalism is true, we should reject Epiphenomenalism.
- (9) If Epiphenomenalism is not true, then we should reject Epiphenomenalism.
- (10) Either Epiphenomenalism is true or it is not true.
- (11) Therefore, we should reject Epiphenomenalism.

In chapter four, I tried to show that this argument was invalid. However, now I will argue simply for the first premise and attempt to show that it is incompatible with the fact of our knowledge of our own qualia. If Epiphenomenalism is true, then we can no longer take our conscious experiences, our pains, to be immediately or directly known. We cannot be

acquainted with causally irrelevant entities. To be acquainted with something, I will assume for the moment, is for the entity one is acquainted with to be the immediate cause of one's belief, or whatever mental state we have when we are acquainted with an entity. To know something immediately is to know it without any causal or inferential intermediaries. If Epiphenomenalism is true, then our qualia cannot be the immediate causes of our beliefs. Thus, we cannot know our qualia immediately.

The basic argument, put somewhat more formally in terms of direct acquaintance with qualia, is as follows.

- (12) For any person S and any x, such that x is S's quale, if S is directly acquainted with x, then x must be the immediate cause of S's belief.
- (13) If Epiphenomenalism is true, then it is not possible that, for any person S and any quale x, such that x is S's quale, x is the immediate cause of S's belief.
- (14) Therefore, if Epiphenomenalism is true, then it is not possible that, for any person S and any quale x, such that x is S's quale, S is directly acquainted with x.
- (15) For any person S and any quale x, such that x is S's quale, S is directly acquainted with x.
- (16) $\therefore$ Epiphenomenalism is false.

This argument begins from the assumption (12) that direct acquaintance requires the object one is acquainted with be the immediate cause of one's internal state. This premise is the one most likely to be contested. (13) is, I take it, a direct result of Epiphenomenalism⁵, epiphenomena since they can cause nothing cannot be immediate causes of anything. The argument is valid. Given the stipulation about the direct acquaintance, the concept of Epiphenomenalism itself and the assumption that we are acquainted directly with our minds, the conclusion follows.

⁵ Note that this argument does not specify Token-E or Type-E. The argument is equally effective against either if properties are the sorts of things we can be acquainted with. I will discuss how these arguments affect the different Epiphenomenalisms in section 6.7.

Again the Epiphenomenalist can insist this argument begs the question by assuming a causal theory of knowledge. The Epiphenomenalist might argue that a subject could know his or her own qualia directly in a non-causal way, perhaps by means of a kind of direct intellectual perception like the direct perception of the Forms imagined by Plato. This model for perception of epiphenomenal qualia *may* be coherent, but I simply cannot imagine what it means to say that one can directly perceive non-causally. Plato never managed anything more convincing than visual, and therefore causal, metaphors for this perception, and I see no reason the Epiphenomenalist could give a better explanation.

The Epiphenomenalist can better respond to problems of direct knowledge of our own minds without insisting on any mysterious direct perception. She could claim that one perceives qualia by means of some mediating physical state that is either nomically or through a common cause to the mind⁶.

Either of these possibilities requires that the Epiphenomenalist abandon the notion that our knowledge of our own quale is direct in the sense that the quale is the immediate cause of our knowledge. However, one might take the notion of direct knowledge to be merely non-inferential knowledge. Thus, we could be said to know directly that the sky is blue or to know directly that one is sitting and typing. These examples seem good candidates for direct, albeit defeasible, knowledge. Candidates like these statements about

⁶ Goldman (1967) in presenting his causal theory of knowledge was forced to deal with the problem of knowing the future. It seemed to Goldman that one can know what will occur, but one cannot have one's beliefs caused by something that has not occurred yet. He solved the problem by appealing to a common cause. For example, some event causes both my belief that I will visit New York and my getting on the plane and arriving in New York.

One might try to generalize the solution to the Epiphenomenalist problem of how we know our minds by appealing to the model of knowing the future on any theory of knowledge and, thus, avoid commitment to a causal theory. The only possible models of knowing the future however are one of the three possibilities mentioned above or an inferential model. An inferential model of knowing the future is unlikely to help the Epiphenomenalist with qualia since inference is not likely to capture the way we seem to know our sensations and perceptions.

the external world have been considered to be direct in foundationalist epistemology yet are clearly mediated by one's senses. The Epiphenomenalist can appeal to a natural sense of direct knowledge as non-inferential knowledge, knowledge that is not immediate but arrived at without inference.

On this revised notion of direct acquaintance or knowledge, Epiphenomenalism appears to be compatible with acquaintance or knowledge of our own minds. One directly knows that one is in pain because one comes to believe that one is in pain without an inference to that belief. This belief may count as knowledge in virtue of its resulting from a common cause or in virtue of the mechanism reliably giving true beliefs or whatever else is required for direct knowledge.

The Epiphenomenalist can respond to the challenge that his theory cannot allow for acquaintance with or direct knowledge of one's own qualia by assuming that acquaintance or direct knowledge do not require immediate causation by the qualia. However, the Epiphenomenalist requires some close relation between the qualia and the belief or knowledge state of the cognizer to ground this acquaintance or knowledge. Presumably this relation is some variety of causal relation or a nomic correlation⁷.

I will now raise two further epistemic objections to Epiphenomenalism based on our purported privileged access to our own minds and on claims of the coherence of skeptical hypotheses. Neither of these objections refutes the Epiphenomenalist provided she appeals to some causal or nomic relation between qualia and the physical and accepts the possibility of defeasibility of our knowledge of our own qualia.

6.3.1 Privileged Access and Skepticism

⁷ This claim assumes that a causal or nomic relation is necessary for a reliable, non-inferential belief-forming mechanism. I assume that a reliable mechanism is more than merely one that accidentally leads to

Relying on a mediated nomic or causal relation may not satisfy many philosophers who think that our knowledge of our own qualia is privileged in a way that knowledge of physical states cannot be. Philosophers with Cartesian intuitions would think that if the theory entails that knowledge of qualia is no more certain than our knowledge of external objects, then some part of our theory must have gone badly awry. If Epiphenomenalism leads to the consequence that the mental is not known incorrigibly, then many philosophers would find that adequate reason for rejecting Epiphenomenalism.

Given the Epiphenomenalist requirement of a physical causal or nomic mechanism. Incorrigibility theorists will admit that it is logically possible for the belief that one's C-fibers are firing to occur without the C-fibers firing. Any relation between a belief and a physical has an inherent possibility of error or unreliability. Any physical mechanism that normally has a certain effect can fail to have that effect. In fact, it is this fallibility of all physical relations that is supposed to make qualia unique. The Epiphenomenalist solution to the problem of how one could have epiphenomenal qualia is that there must be some mediating physical state between one's belief and the quale itself. If such a mediating physical state is required, and if the connection between that mediating physical state and the belief can be broken, then the connection between the belief and the quale must be broken as well. Since it is at least logically possible that the mediating physical state occur when the belief does not, it is also possible that the belief might occur without the quale. And for this to be possible for a belief and a quale shows that the belief is not incorrigible. Thus, the Epiphenomenalism is inconsistent with incorrigible access to one's own qualia.

Nonetheless, a weaker form of privileged access to one's qualia can be maintained. One may know the quale better than one knows the physical intermediaries. In fact, one

true beliefs.

need not know the physical intermediaries at all to know the quale if Epiphenomenalism is true. Unless we are cognitive scientists, we have no idea, for example, what process converts a two-dimensional set of dots on the retina into a three dimensional picture of the world; yet, the process of turning that two dimensional image into a three dimensional picture causally mediates our perception of the world. Clearly, then, we need not know the processes that lead to a belief for us to come to have that belief. *A fortiori*, we need not know the causal intermediaries any better than we know the results. Thus, Epiphenomenalism does not require that we do not in general have better access to our minds than to the causal (or nomological) intermediaries between our beliefs and our qualia. Although Epiphenomenalism is inconsistent with incorrigibility, the Epiphenomenalist may still think that qualia are known as well as objects of perception.

Rejecting Cartesian intuitions may create some tension with the reasons many philosophers have for accepting property dualism. Chalmers (1996), for example, thinks that taking consciousness seriously requires that we believe the mental is known better than the physical world. If one is the sort of person who accepts Epiphenomenalism, then one is likely to think that our mental states have a certain privileged status that other states do not. If one thinks that our mental states are known only as well as physical objects, say perceptual objects, then one has taken away much of the reason many philosophers have for thinking the mental is unique, that it resists reduction to the merely physical. One of the common reasons for thinking the mental is irreducible is its unique or privileged epistemic status. Thus, giving up the mental's unique epistemic status is antithetical to some motivations for Epiphenomenalism.

If Epiphenomenalism is correct, then we have to give up the notion that the mental is better known than perceptions of the physical world. If qualia are correlated with physical states that give rise to our beliefs, then we could be wrong about the qualia just as we could be wrong about our perceptions of the external world.

A second objection along these lines is that Epiphenomenalism cannot make sense of standard skeptical problems⁸. If Epiphenomenalism is true, we cannot even formulate skeptical problems, the argument goes, that philosophers have thought we could. Skeptical arguments require that we could be appeared to just as we are even if the external world did not exist or even if everything we believe about it were wrong. Epiphenomenalism must say that we cannot be appeared to as we are without any physical state existing. We could have skeptical doubts about the nature of the external world but not with the existence about something physical. So puzzles about the existence of a world outside our minds cannot even arise if Epiphenomenalism is true.

The Epiphenomenalist response is that belief in Epiphenomenalism does not meet the high standards of proof demanded by the skeptic. It is logically possible, or at least conceivable, that Epiphenomenalism is false, and that is all that is required to formulate the skeptical problem. This fact about skepticism is irrelevant to Epiphenomenalism. For example, if we were to accept the above argument we would have to admit that if the type identity theory is true, then we cannot state skepticism properly because that view claims that our qualia and brain states are identical. But this fact about skepticism, presumably,

⁸ Chalmers (1994) suggests an argument along these lines, but argues that the close access we have to our own consciousness allows skeptical problems to occur even if Epiphenomenalism is true. Chalmers thinks that taking consciousness seriously, i.e. eschewing eliminativism and reductionism about consciousness, requires that we have epistemic access to our experiences although this access may not involve beliefs about them. This access, on his view, is closer than one could have with a causally mediated physical state, so he thinks rejecting the argument from skepticism is a natural consequence of taking consciousness seriously. What this epistemic access might consist in, I find as wholly mysterious as the possible direct

does not provide a reason to think the identity theory is false. It merely shows that the identity theory does not itself live up to the standards the skeptic demands for knowledge. The same claim is true of Epiphenomenalism. Epiphenomenalism does not live up to the strict demands for knowledge that skepticism requires. So all this criticism shows is that skepticism is a problem for Epiphenomenalism just as it is for any theory that asserts anything. Epiphenomenalism has no special problem with skepticism.

Epiphenomenalism conflicts with incorrigibility of knowledge of our qualia but is compatible with fallible but noninferential access to our qualia. Skeptical problems are reasonably statable even if Epiphenomenalism is true. And, provided there is a nomic connection of some sort between the quale and a physical state, we seem able to know or form beliefs about our own minds, refer to them and represent them.

6.3.2 Acquaintance with Qualia

If Epiphenomenalism is correct, then it appears to be impossible for us to be acquainted with our own qualia. I will argue in this part that the Epiphenomenalist can defend his position by assuming a close nomic or causal connection between qualia and physical or brain states.

In chapter four I argued that Epiphenomenalism is coherent by appeal to the possibility of knowing propositions about fictional and nonexistent entities. This model may work for knowing propositions about epiphenomenal qualia, which was enough to defeat the incoherence arguments of chapter four, but this model of knowing qualia is insufficient for acquaintance with those qualia. And, I assume, we are acquainted with our qualia. Thus, the Epiphenomenalist cannot pursue an analogy between qualia and fictional entities to deal with the problem of first-person acquaintance with qualia.

perception I have already dismissed.

A brief digression on qualia realism, or reification, is in order here. One theory of qualia that competes with this reifying qualia is the adverbial model of qualia. On the adverbial model of qualia, qualia are not entities with which we are acquainted but are ways in which we perceive or are appeared to. So, when I see red, I am not acquainted with a red percept or quale but I perceive *redly*, I am in a redly-sensing state. This view would deny that we are acquainted with qualia at all. In fact, we would not, strictly speaking, know qualia at all, since qualia could not be *objects* of knowledge on this view. This adverbial model may be correct, but it does not affect my argument in any deep way. If the adverbial theory is correct, then we would need to rephrase the argument, but that is all. The adverbialist would admit that we are appeared to in certain qualitative ways, and presumably he would admit that these appearances are known to us. So, I merely have to argue that there must be some causal relevance of these adverbs, these properties of our experiences, for us to be directly aware of the way we are appeared to. And any view would have to admit that we are directly aware, defeasibly, of how we are appeared to. That is all my argument really requires, that we be directly aware of how we are appeared to. Whether we are aware of an object or only of a property of our experience, these qualia are still part of our direct conscious experience. Since it is convenient, I will continue to speak as if qualia are objects, but the argument should apply to the adverbial view as well.

It is possible that we know epiphenomenal qualia in the same way we know nonexistent or fictional entities, but this model will not allow us to be acquainted with qualia. The distinction between acquaintance and propositional knowledge played some underlying role in my previous argument that Epiphenomenalism is coherent. It is important to make the distinction here to show that we can have knowledge but still not acquaintance. Having

propositional knowledge of epiphenomenal qualia is not problematic. We can have propositional knowledge of fictional and nonexistent entities; by analogy, we should be able to have knowledge of epiphenomenal qualia. We know that Sherlock Holmes was a brilliant detective who noticed details that escaped everyone else. We know that unicorns do not exist. We may know, or at least believe, that epiphenomenal qualia must be unusual or odd phenomena. Perhaps we can have propositional knowledge of qualia in whatever way we have such knowledge of nonexistent and fictional entities. If we can have propositional knowledge in this way, then this knowledge may not require a nomic or causal connection. We have no causal connection to Sherlock Holmes or to unicorns since none of them exist, nor can we have a nomic connection to nonexistent entities except in some trivial sense. So, if the model of knowledge of nonexistent and fictional entities holds for propositional knowledge of qualia, we need not appeal to a nomic or causal connection between our beliefs and the qualia.

Some philosophers (e.g. Lycan 1987) may be tempted by this approach to knowledge of qualia. Certainly this approach will not appeal to realists about qualia (e.g. Jackson 1982 and Chalmers 1996), and, moreover, this approach will not allow for acquaintance with qualia that I am supposing we have.

However, acquaintance and knowledge are quite distinct. We know facts about Holmes, but we have no acquaintance with him. Appealing to the model of knowledge of fictional and nonexistent objects solved our problem for the Argument against the Coherence of Epiphenomenalism. But this model cannot help with the problem of acquaintance with our own minds since we are not acquainted with fictional or nonexistent entities. The best Epiphenomenalist response is, again, to appeal to a close relation between the qualia and the

physical states by means of which a person is acquainted with the qualia or is aware of the quality of his experiences.

6.4 Problems with (3): Qualia as Representations

The third response to the rejection of qualia Epiphenomenalism (3) raises additional problems for Epiphenomenalism. The response (3) is the claim that qualia are the means by which we perceive the world; qualia represent the world to a subject. This view need not be the controversial idea that qualia are pictures or objects of perception by some inner eye. All this view claims is that qualia represent states of affairs, external states of affairs like the fact that the stove is hot, or internal states of affairs, like the fact that my stomach is empty. But in whatever way qualia represent, they must represent to a subject. However, if Epiphenomenalism is true, then it appears to be impossible for qualia to do this. If qualia are causally irrelevant, then they cannot represent anything to anyone.

One view of representation, perhaps a simplistic view, is that representations must represent to a subject. The representation must be causally relevant to the subject in some way, perhaps not as a simple percept, but it must somehow relate to the subject. It is hard to see what kind of relation this could be other than a causal or nomic one. So, the possibility that qualia represent the world seems to require that they be causally relevant to a subjects beliefs and other internal states. It is impossible for epiphenomena to fulfill this role, so either they are not representations, contra this response, or they are causally related by means of some mediating physical states.

Mental representation is a large and difficult issue in contemporary philosophy of mind. Perhaps the idea that representations must represent to a subject reduces to a rebarbative Cartesian Theater (Dennett 1994 and elsewhere) view, an unnecessary holdover

from a substance dualism that sees the self as an homunculus viewing internal pictures. However, my point can be made without making this kind of controversial assumptions about the nature of this representation. Most theories of representation require that an internal token represents because that token is related in some, perhaps complex, causal, counterfactual or nomic way with the fact that it represents. Any theory along these lines is, so far, compatible with Epiphenomenalism. However, any theory of representation must commit to the view that the representation is had, stored by, or acted upon by a subject. For example, for my token of a syntactic representation, #water#, to represent water requires that the token occur in a system with a complex structure with inputs and outputs relating to the world. On most theories this syntactic form #water# encodes some information about the world by means of a nomic, causal or counterfactual relation to the objects in the world, either internal or external to the system. For this token to be part of such a complex system, it must be causally related to the other parts of the system. Epiphenomenal qualia cannot be a cause in any such system, so they must be effects of some cause that does play such a role.

Barring an idiosyncratic theory of representation, the Qualia Epiphenomenalist's view that qualia are representations commits the Epiphenomenalist to the view that qualia are nomically or causally connected to some physical state or token that plays a causal role in the organism's internal information processing system.

6.5 Problems with Reference to Qualia

The last problem I will raise for Epiphenomenalism is one of reference to one's own qualia. I will argue that this problem of reference to one's own qualia can only be solved by assuming a close nomic, causal or metaphysically necessary connection between a physical property and the qualia itself.

The argument against Epiphenomenalism from the problem of reference is that since we cannot refer to epiphenomena directly, we cannot rigidly designate them at all. The argument in outline form is as follows.

- (17) If Epiphenomenalism is true, then it is not possible that for any person S and any quale x, such that x is S's quale, S rigidly designates x.
- (18) It is possible that for any person S and any x such that x is S's quale, S rigidly designates x..
- (19) $\therefore$ Epiphenomenalism is false.

According to this argument, Epiphenomenalism does not allow one to rigidly designate one's qualia. Since one cannot be directly causally related to one's qualia, one cannot rigidly designate them. The case is analogous to that of Newman⁹. Imagine that we invent a name, say "Newman₂" to refer to the first person born in the 22nd century. This name does not refer to the same person in all possible worlds since the name may apply to one of two possible individuals, say Sam and Terry, either of whom might be born before the other. Thus, "Newman₂" cannot refer to either individual rigidly.

To designate an individual rigidly is to pick out that individual in all possible worlds. Thus, the term "Bill Clinton" picks out *that person*, Bill Clinton in all possible worlds. It is impossible for Bill Clinton not to be Bill Clinton, so the term identifies Bill Clinton in all possible worlds. On the other hand, the term, "the second President to be impeached," does not pick out Bill Clinton in all possible worlds. Several possible scenarios would falsify the statement "Bill Clinton is the second President to be impeached." First, Clinton might not have been impeached. Second, Andrew Johnson might not have been impeached, so Clinton would have been the first President to be impeached. Third, Richard Nixon or some other President might have been impeached, so Clinton would have been at least the third President to be impeached. Definite descriptions that are contingently true of the objects

they designate, then, do not rigidly designate those objects. Names and natural kind terms, like "pain," by contrast, designate objects rigidly.

In the case of Newman₂ we have to say that "Newman₂" does not pick out one person across all possible worlds because more than one person might have been born first in the 22nd century. One possible response to this puzzle is to claim that "Newman₂" is a non-rigid designator, like a definite description, that picks out a person in the actual world but which would not identify the person across all possible worlds. We should not accept this response, however, because it treats "Newman₂" as something other than a name, perhaps a disguised definite description. But "Newman₂" certainly appears to be a name. Treating some apparent names as something other than names raises problems for determining when a name is really a name and when it is a disguised definite description. If we could somehow distinguish these two types of "names," then we would have to resort to a bifurcated theory of names, one theory for one type of names, for example "Newman₂," and a second theory for another type, for example, "Bill Clinton." But the two types of names and "Bill Clinton," for all other purposes must be treated as the same kind of thing. So treating names as only one type of thing avoids these problems.

The better option in this case is to treat the term "Newman₂" as a name that does not yet designate at all. Since no one has yet been born in the 22nd century, the name does not refer to anyone yet. This example shows that in general we cannot refer, with a rigid designator, to someone who does not yet exist.

This result affects Epiphenomenalism, first, in that we cannot take reference to minds to work on the model of reference to the future. Since future events could turn out differently, we cannot generally use terms to rigidly designate future entities. If we are to

⁹ This case is discussed in Adams (1986).

refer to our minds, then we must be able to refer to them with rigid designators. For example, pains are generally taken to be a natural kind, but natural kind terms rigidly designate their referents. If that's correct, then when we attempt to refer to our pains, we must fail to designate anything at all. Unless we take all mental terms to be disguised descriptions or some other nonrigid designator, which I think anyone would find implausible, then we cannot both accept a model of reference for the mind following a model of reference to the future and believe that we can refer to our minds.

One exception to the rule that we cannot refer using rigid designators to future entities is the case in which a present object determines some future object in a metaphysically necessary way. For example, following Kripke (1972), I can rigidly designate a future person by referring to the union of *that* sperm and *that* egg, presuming that the union could only result in *that* person. Metaphysical necessity, on Kripke's view, holds across all possible worlds, in the way that identity does, but it implies nothing about the analyticity of such relations that broadly logical possibility does. So, on Kripke's view, that Hesperus is Phosphorus is metaphysically necessary. Both terms refer to the same object rigidly, and that object, the planet Venus, could not fail to be Venus. But this kind of necessity is not obviously logical necessity since there is no logical requirement that any identity claim involving two different terms must be true. Thus, we can rigidly designate some future entities provided that the present entities metaphysically necessitate the future ones.

This exception provides the Epiphenomenalist's best response. We can refer, the Epiphenomenalist would claim, to our own minds because they are necessitated strongly by the physical entities underlying them. Following Kripke's suggestion, we might think that

the necessitation must be metaphysical, but perhaps a weaker nomological necessity would suffice.

Moreover, a purely nomological relation between the referring term and the entity designated may be sufficient for reference. For example, in a dubbing ceremony, or in whatever way a term first gets fixed to the person or entity, the naming requires only a nomological or causal connection between the entity dubbed and the term that refers to it. For example, when one refers to cats by ostension, one points to a cat or set of cats and says, effectively, the term “cats” refers to that kind of thing. One’s act of ostension is mediated by causal or nomic intermediaries. Thus, acts of ostension, for example, which are perfectly adequate for fixing a rigid designation, require only certain kinds of causal or nomic relations. So, reference to epiphenomena may only require a nomological or causal relation between the term and the object.

I have shown that the Epiphenomenalist’s best, perhaps only, response to a series of problems involving knowledge, belief formation, representation and reference is to postulate a close nomic or causal relation between the quale and some underlying physical state. I will now turn to a final argument that counts against Epiphenomenalism. The final argument I will present is that there is no reason to think the nomic connection is of the right kind to support knowledge, belief formation, reference or representation.

6.6 Objection to the Nomic Correlation Response

The nomic or causal connection between the quale and the physical entity is not the right kind of nomic connection to ground acquaintance or reference. Some but not all nomic or causal connections can provide this grounding. For instance, the sputtering noise my car makes is nomically related to its lacking a belt; whenever the belt is missing, as a matter of

law, the car makes a sputtering noise. The missing belt is also causally related to the sputtering noise; the lack of a belt causes the sputtering noise¹⁰. This sputtering noise is nomically related to my beliefs about the noise, and it causes me to have worries about the noise. Thus, there are nomic and causal relations between my beliefs and the sputtering noise and between my beliefs and the missing belt. However, intuitively, I am acquainted only with the sputtering noise not the missing belt. Moreover, I do not refer to the missing belt when I say, "There's that sputtering noise again." The nomic or causal relations between the sputtering noise and my beliefs and utterances ground knowledge and reference; the nomic and causal relations between the missing belt and my beliefs and utterances simply are not the right kind of relations to ground knowledge and reference¹¹. Some nomic or causal connections can ground these relations whereas other such connections cannot.

The Epiphenomenalist can give no reason the connection between qualia and knowledge or utterance is the correct kind without begging the question. The cases to which the Epiphenomenalist must appeal are cases of perception or ostension in which the object perceived or ostended is ultimately a cause of the belief or utterance. These acts of perception or ostension are mediated, but the thing perceived or ostended is still a mediate cause (or perhaps more accurately one of a set of causal conditions) of the belief or act of referring. So, given that not all nomic or causal connections, can ground such relations, the Epiphenomenalist needs more than these loose analogies. The analogies appealed to appear

¹⁰ How can a lack or an absence be a cause? I don't know, but it seems to be true that absences can be causes. The absence of oxygen in the brain causes humans to fall unconscious and ultimately die. A lack of careful oversight may be a partial cause of industrial accidents. Companies are often held responsible not for what they have done wrong, but for what they have failed to do. The notion of an absence being a cause is puzzling but common, and in any case not relevant here. Another example could be found if this one is too puzzling.

¹¹ Ultimately, given the Exclusion Argument, we may not be able to know or refer to the sputtering noise either, but this conclusion, which we can avoid if we find something wrong with the Exclusion Argument, does not affect the example as a useful illustration of the intuition.

not to be relevantly similar to the case of acquaintance with and reference to epiphenomenal qualia.¹²

My argument here is not the fallacy that because these two things, the sputtering and the missing belt, are not both known to me, they are not identical. The argument is just that, given two distinct entities that are both nomically related in some way to my beliefs, only some of them are related in the correct way to ground knowledge. So, given that the entities are distinct, even though both have some nomic or causal relation to me, I do not necessarily know or refer to both.

The Epiphenomenalist may insist that, nonetheless, the relation of the quale to these states is one that can ground such a relation, but she cannot provide an independent reason (aside from a need to avoid the problem of mental causation) for thinking this claim is true. It is hard to see how one might specify the right kinds of nomic connections without simply begging the question one way or the other. However, given the disanalogy between the cases the Epiphenomenalist needs to appeal to as a model, the burden of proof should be on the Epiphenomenalist to show that epiphenomenal qualia are related in the correct way.

My main worry about this line of argument is that it may beg the question. I do not want to assume a simple causal theory of knowledge and reference, yet, when the possibility of knowledge of and reference to an entity that does not fit the simple causal theory, I claim that it needs to mimic the simple causal theory. Thus, my argument appears to make a *de facto* assumption of a simple causal theory.

¹² Perhaps a strong metaphysically necessary relation between the epiphenomenal qualia and the belief may ground knowledge or reference as in Kripke's sperm and egg designation. But the same problem can be raised for this view, I think. The relation between a new person, an effect of common causes that metaphysically necessitate that person's existence, may be analogous to the relation of qualia to the physical, but it seems in these cases that we have to know that the necessary connection holds for the attempted reference to hold. Whether such knowledge is possible is still very much at issue.

My response to this worry is that it is incorrect. There are reasonable exceptions to these causal theories. Abstract, fictional, and nonexistent entities all appear to be exceptions to simple causal theories of knowledge and reference. That we can know and refer to some or all of these entities shows that a simple causal theory is not correct, or at least not comprehensive. However, as I have noted, epiphenomenal qualia do not fit any of these models. Thus, the Epiphenomenalist needs to give some reason to expand the class of entities we can know and refer to include epiphenomena, and I do not see how she could accomplish this task. My argument, then, does not assume a de facto simple causal theory, but merely demands a reason to include a problematic new category to those categories that already provide exceptions to the simple causal theory.

6.7 Type-E and Token-E Revisited

One point that needs to be addressed is whether these arguments are convincing against type Epiphenomenalism. Type-E, recall, is the view that the events do not occur in virtue of falling under a mental type or in virtue of their mental properties. The following propositions restate the argument in terms of properties.

- (20) If Type-E is true, we can be neither acquainted with nor refer to the conscious properties, the qualitative character of the experiences, of physical events.
- (21) We can be acquainted with or refer to our own conscious properties.
- (22) $\therefore$ Type-E is false.

If we are acquainted with or refer to properties of objects as well as objects themselves, then the argument of this chapter affects Type-E just as it does Token-E. If we can be acquainted with and refer to the qualitative aspects of our experiences, treating qualia as properties, then those properties must be related in the right nomic or causal way. But they are not related in the right way. Thus, Type-E is shown to be false if Token-E is.

I think that the case of Type-E is entirely analogous to the case of Token-E. We do seem to be acquainted with the painfulness of our experiences, the pleasurable-ness of our sensations, and, in general, the properties of our mental events and states. At least, we are directly aware of these properties or qualities of our experiences. We are able to talk about them and designate them rigidly. Thus, if Type-E is taken as a claim about events falling under types rather than as a claim about properties, perhaps from a suspicion about properties, I can put the argument in less contentious terms.

(23) If Type-E is true, we cannot be acquainted with or refer to our qualia insofar as they are mental or conscious.

(24) We can be acquainted with or refer to our own qualia insofar as they are mental or conscious.

(25) $\therefore$ Type-E is false.

Whatever one's favorite phrase for describing Type-E, the argument can be given for that phrasing. It seems uncontroversial that we can be acquainted with mental events in virtue of their falling under a mental type. In fact, most of our acquaintance with our own minds is in virtue of the event falling under a mental type. It does no good to object that we are acquainted with objects only and not properties because we still are acquainted with them as mental. We can still refer to our mental events or states *qua* conscious events. To deny this claim is to deny a fact as obvious as that we are acquainted with our own minds at all, but we could not be acquainted or refer if the nomic or causal relation between the qualia and the physical state is not of the right sort. Thus, Type-E does not avoid the argument against Epiphenomenalism as I have framed it in this chapter.

6.8 Conclusion

I have argued in this chapter that Qualia Epiphenomenalism must be false. Causal or nomic correlations are not enough to ground knowledge and reference, and we have no good

reason to think that the relations of epiphenomenal qualia to our beliefs and acts of referring are of this correct sort. This argument is not a knock-down argument against Epiphenomenalism. It is possible that the nomic or causal relation involved in Epiphenomenalism really is the right kind to ground these relations. However, this argument does show that Epiphenomenalism is most likely false. Epiphenomenalism is not likely to be the best available option in solving the problem of mental causation.

I have some qualms about arguing *a priori* that the mental must have causal power. Ordinarily we discover from experience whether some event is a cause or some property is causally relevant to the production of an effect; *a priori* considerations should not in general tell us whether a property is causally relevant to an effect. However, *a priori* considerations might tell us whether it is possible for some property to be completely causally irrelevant. In addition, the argument for the causal relevance of the mind does rely on some empirical evidence. The argument relies, for example, on the facts that we know and can refer to our own minds. These claims, I suppose, could be shown empirically to be false although the evidence for them seems to be so overwhelming as to need no argument.

The most promising solution, then, to the problem of mental causation, as it is presented by the Exclusion Argument, is to find this argument unsound, to find reason to reject one of its premises or to show that the conclusion does not follow from the premises. Two of the premises of the Exclusion Argument are especially controversial. Thus, the goal of the remaining part of this work, chapters seven and eight, is to consider whether either of these controversial premises of the Exclusion Argument might be false, and false in a way that allows for mental causation. So, in chapter seven, I will argue that, although the supervenience thesis is almost certainly false, that fact provides no mechanism for the

mental to be causally relevant, and so cannot help in solving the problem of mental causation. In chapter eight I will argue that the Exclusion Principle is false for closely related sets of properties, and so the mental can be causally relevant if it is closely related to the physical.

PART III

PREMISES OF THE EXCLUSION ARGUMENT

OVERVIEW

In previous chapters I have carefully analyzed the Exclusion Argument and the problem it raises for the possibility of mental causation. In the previous part of this work, chapters four through six, I argued that Epiphenomenalism could not be correct. If Epiphenomenalism is not correct, then the Exclusion Argument must be unsound. It is now time to investigate the validity of the Exclusion Argument and the truth of its premises.

I have characterized the Exclusion Argument in its most general physicalist form as follows.

1. *Dualism*. The mental and the physical are real.
2. *Irreducibility*. The mental and physical are not type identical.
3. *Supervenience*. The mental supervenes on or is determined by the physical.
4. *Causal Closure of the Physical*. No physical effect can occur without a physical cause or except in virtue of a physical property of its cause.
5. *Generative/Determinative Exclusion Principle*. For any effect or property, there can be only one complete, independent cause or set of causally relevant properties.
6. Mental causation is not a case of causal overdetermination.
7. *Conclusion: Epiphenomenalism*. The mental is not causally relevant to the production of any effect.

These premises lead to the conclusion that the mental is epiphenomenal in the following way. The mental and physical are real entities that are not type-identical (1) and (2). The physical can only be caused by something physical (4), so there is no independent downwards causation from the mental to the physical. The mental is determined by the physical (3), so the mental effect is determined by the physical. Thus, the physical determines all the effects there are, both mental and physical, and, since

when there is one complete, independent cause, there cannot be another (5), the mental must be causally irrelevant.

I argued in the previous part of this work that we cannot accept Epiphenomenalism, the conclusion of the Exclusion Argument. It follows from this that either at least one of the premises of the Exclusion Argument is false or the argument is invalid. If a premise is false, importantly, it must be false in a way that allows for mental causation.

The most frequently questioned of the premises are the *Irreducibility* of the mental to the physical, the *Supervenience* of the mental on the physical, and the *Exclusion Principle* itself. The irreducibility thesis has been criticized and defended at length elsewhere¹. At least some of that criticism, and the most effective to my mind, arises directly from the Exclusion Argument² and worries about mental causation. Part of the purpose of this work is to defend Nonreductive Physicalism against this reductionist attack. If I can defend Nonreductive Physicalism, which assumes Irreducibility, against the problem of mental causation, then there is no need to argue further for the Irreducibility thesis. Thus, I will not discuss the irreducibility of the mental to the physical any further.

The Causal Closure of the Physical has only infrequently been defended. Kim (1996 and elsewhere) defends it by claiming that it is a basic presupposition of Physicalism, that to deny it is to deny the completeness of physics and to postulate mysterious entelechies and downwards causation. Further, he argues that the in-principle incompleteness of physics makes it impossible to study physics alone without including

¹ Critics of irreducibility include Bickle (1992), Churchland (1984), Kim (1992 and 1993b), and Searle (1992). Defenders include Fodor (1974), Pereboom and Kornblith (1991).

some element of psychology or other higher-order discipline. On the other hand, some in the popular media have sought to deny the Causal Closure of the Physical, but the mechanisms they have postulated are lacking in plausibility. Although a more detailed discussion of these issues is warranted, I will not address the question here. I will, instead, focus on two premises that are more controversial.

Philosophers have given reason to doubt two other premises, the Supervenience thesis and the Exclusion Principle. In the chapters that follow, I will address the possibility that these principles might be false, and see if their falsity can allow for some mechanism by means of which the mental might be causally relevant.

² I have in mind Kim (1993b).

CHAPTER 7

SUPERVENIENCE AND MENTAL CAUSATION

7.1 Introduction

In this chapter I will examine whether the falsity of the Supervenience Thesis can allow for mental causation. I will explore some reasons that one might give for thinking mental or content properties do not supervene on physical properties. Then I will consider whether any of these reasons for failure of supervenience allows for mental causation. My thesis is that, although the Supervenience Thesis is almost certainly false, at least for content properties, this failure of supervenience does not allow for content, or the mental, to be causally relevant.

Probably the most common and appealing response to the Exclusion Argument is to argue for a dual-explanandum³ view. A dual-explanandum view of mental explanation is that the mental and physical are two different domains requiring separate explanations. Behavior and some mental effects can be caused and explained by the mental whereas mere bodily movements and physical effects require only physical causes and explanations. The dual-explanandum theory's postulation of two separate sets of explanations, and given explanatory realism, the dualism that entails, fits nicely with our commonsense and scientific explanatory practice. And dual-explanandum theories, as they surely ought, allow our mental and special science explanations to remain as real and scientifically respectable as physics explanations. Moreover, these theories do not buy mental causation by postulating any mysterious downwards causation. Of the

³ "Dual-explanandum," rather than "dual-explananda" or "dual explanans" is Kim's (1991) preferred term. The choice of nomenclature does not matter. I will follow his usage merely for consistency.

possible solutions to the problem of mental causation, dual-explanandum theories best fit our scientific practice and general physicalist commitments.

However, dual-explanandum views founder on the two principles of Supervenience and Exclusion. If the mental supervenes on the physical, then the mental is determined by the physical, and so there is no need to postulate this extra causal power for the mental. There may be two explananda, but if supervenience obtains, then there is only one explanans. And, according to the Exclusion Principle, because there is already one cause (or explanation) there cannot be another. So, dual-explanandum theories, although they meet our intuitions best of the possible theories of mental causation, are incompatible with these two premises of the Exclusion Argument. So, for one to accept a dual-explanandum theory, one must reject at least one of these premises.

As we noted in chapters two and three, there are several different varieties of or ways of formulating supervenience. The three most common types of supervenience are weak, strong and global supervenience, each formulated as relations between sets of properties. Weak and strong supervenience are both construed as local supervenience claims. Kim formulates these claims as follows.

Weak supervenience:

Necessarily, for any object *x* and any property *F* in *A*, if *x* has *F*, then there exists a property *G* in *B* such that *x* has *G*, and if any *y* has *G* it has *F*. (1987b, reprinted 1993a, 80)

Strong supervenience:

Necessarily, for any object *x* and any property *F* in *A*, if *x* has *F*, then there exists a property *G* in *B* such that *x* has *G*, and necessarily if any *y* has *G* it has *F*. (1987b, reprinted 1993, 80)

Global supervenience:

Any two worlds indiscernible with respect to B-properties are indiscernible with respect to A-properties. (1987b, reprinted 1993a, 82)

I argued in chapters two and three that strong supervenience was closest to capturing our intuitive notion of the dependence of the mental on the physical. I also argued that, given the common belief that causal relevance requires necessitation of effects, strong supervenience is necessary for the Exclusion Argument to be valid. Thus, when I discuss failure of supervenience, I will generally mean failure of strong supervenience. However, the first two reasons I give for failure of supervenience will affect all of the supervenience claims. The third reason I give will affect only the first two, weak and strong, but will still allow for global supervenience to obtain.

The failure of mental content to supervene provides an intuitive way of overcoming this problem for a dual-explanandum view. Failure of supervenience seems to allow for an autonomous domain, behavior or the mental, which is not determined, or not fully determined, by the physical. Thus, the mental could cause mental effects or behavior, and the physical could cause physical effects and pure bodily movements. If these two domains are autonomous, then the Exclusion Argument could not exclude either from causal relevance. I will conclude that even though it is likely that the mental or mental content does not supervene on the physical, no mechanism is adequate for the mental to be causally relevant.

I will give three reasons one might reject the supervenience thesis. I will explain how each of these reasons might support a dual-explanandum theory. Finally, I will show that, for each of these reasons, the dual-explanandum theory cannot provide an adequate theory of mental causation.

7.2 Criticism of Supervenience

Several philosophers⁴ suggest that the solution to the problem of mental causation may rest on the failure of the mental to supervene on the physical. Three types of evidence suggest that mental properties do not supervene on physical properties. For any of these reasons for failure of supervenience to allow for mental causation, there must be some dual-explanandum theory that takes advantage of this failure of supervenience. For failure of supervenience to allow for mental causation, there must be two explananda that require separate explanations. I will examine what dual-explanandum theories might appeal to each of the reasons for failure of supervenience.

First, the conceivability of absent or inverted qualia is sometimes thought to entail the logical possibility of failure of phenomenal properties to supervene on the physical. Second, the failure of properties of quantum systems to supervene on properties of parts of these systems may provide some reason to doubt that the mental supervenes on the physical. Third, the role of context in individuation of content and special science properties shows that these properties do not supervene on local physical properties.

7.3. Consciousness

To solve the problem of mental causation each of these possible reasons for rejection of supervenience must support a theory of mental causation. One might appeal to a failure of supervenience for consciousness to support a dual-explanandum view of conscious effects. For example, if my conscious mental states are not determined by the physical, then logical space exists for the mental to cause these conscious mental states. My pain may be causally relevant to my conscious knowledge of it on this view, but my pain does not cause my behavior since my behavior supervenes on the physical actions I

⁴ Baker (1993) rejects supervenience and any other metaphysical commitment. Menzies (1988) rejects what he calls "causal reductionism" or supervenience of causal powers.

make. Jackson (1982), for example, allows for a dual-explanandum view involving mental-to-mental causal relevance because he believes that pure Epiphenomenalism would raise problems of knowing our own minds.

Jackson's dual-explanandum view, as other such views do, conflicts with the Supervenience thesis and the Exclusion Principle. For example, if our knowledge of our own minds is determined by the physical, then there is no need for a mental cause of this knowledge. And if there is already a cause of the knowledge, then there cannot be another, conscious, cause of it. Jackson's view, then, can only work if one of these premises, the Supervenience thesis or the Exclusion Principle, is rejected.

Jackson and other qualia boosters appear to have a ready response to this problem, however. As noted above, the first possible reason to think the mental does not supervene on the physical is that, as some philosophers (e.g. recently Chalmers 1996) have argued, it is conceivable, and therefore logically possible, that phenomenal properties do not supervene on physical properties. Absent qualia (or in Chalmers's terms "qualia zombies") and inverted qualia cases seem conceivable⁵.

It is conceivable that one could have the very same physical or functional states that one actually has but have no correlated conscious, qualitative or phenomenal state. For example, Fred hits his thumb with a hammer; his nervous system sends a signal to his brain; his C-fibers fire; and he feels a sensation of pain in his thumb. The intuition behind absent qualia cases is that it is logically possible that all those physical or functional effects occur to Fred, but Fred feels nothing at all. It seems conceivable⁶ that

⁵ Kripke's (1972) modal argument against type and token identity theories can be seen as a kind of absent qualia argument.

⁶ One might claim that nothing that is logically impossible is really conceivable but only apparently conceivable. The basic point can be made assuming that these cases are only apparently conceivable.

he have all the same physical or functional states but not feel anything at all, or at any rate, his physical or functional states may have no qualitative or subjective feel. These cases are sometimes called “absent qualia” cases because we imagine that one has a physical or functional state but these states lack qualitative content.

Inverted qualia cases are that it is conceivable that a person might have the same physical or functional states as anyone else actually has, but these states have a different qualitative feel for this person. Thus, Fred might have green sensations when looking at red objects and vice versa. So, he might look at a ripe tomato and have the qualitative sensation that Ted has when Ted looks at a something green, say the grass in his well-tended front lawn. Both the absent qualia and the inverted qualia cases claim that qualia are conceivably different even with the same underlying physical state. Qualia, if these cases are genuinely possible, do not supervene on physical or functional states. The claim that conceivability entails logical possibility is, to the say the least, contentious and I do not wish to support or contradict it here, but any philosopher who accepts the logical possibility of phenomenal properties differing without a difference in physical properties would deny that the mental supervenes on the physical.

It is one thing to deny supervenience of qualia on the physical but another to show how this failure of supervenience may allow for mental causation. Supplementing Jackson’s dual-explanandum theory for qualia with a rejection of only the Supervenience thesis entails that the only things that could be caused by the mental would be those things that did not supervene on the physical. This means that, since consciousness fails to supervene, only conscious, qualitative effects could have mental causes. So, Jackson’s dual-explanandum view appears to solve the problem of mental causation by appealing

only to the failure of supervenience of conscious mental events on physical events, a failure that he, or at least other qualia boosters, may accept.

However, there are several problems with this theory. First, this solution depends on the contentious claim that conceivability entails logical possibility. This problem is quite serious, but I will ignore it to avoid going too far afield. Even if this claim is correct, other problems threaten this solution. Second, this solution may not avoid the problems of reference that full-blooded Epiphenomenalism had. Third, since this view appeals to failure of supervenience only for logical necessity but not nomic necessity, it does not help with the problem of mental causation since causation is only a nomic relation. Fourth, this solution does not give the kind of mental causation we want. Fifth, this theory seems to involve a mysterious parallelism between the conscious and the physical or panpsychism. Avoiding both of these problems seems to lead one directly back to the problem of mental causation.

Jackson (1982) postulates a limited Epiphenomenalism, which I am here suggesting is a kind of dual-explanandum view, in which the mental can be causally relevant at least to our beliefs. Jackson thinks causal relevance of the mental is necessary for us to know our own minds. Assuming a failure of supervenience as an auxiliary hypothesis⁷, one might make room for causal power of the mental only if the knowledge or beliefs one has about it are not themselves determined by the physical. If we assume that conscious mental states do not supervene on the physical, then we should think only that conscious mental states can have mental causes. If the other premises of the Exclusion Argument are true, then the only room for causal relevance of our minds is for

⁷ I do not know whether Jackson would commit to this additional hypothesis.

the mind to be causally relevant to events or properties which are not determined by the physical.

For one to accept this solution, based on this proposed dual-explanandum view, one should expect the theory to solve the problems of knowledge of one's own mind that prompted Jackson's limitation on Epiphenomenalism in the first place. If knowledge is a conscious mental state, then this dual-explanandum view does solve that problem. If knowledge is conscious, on this view, it fails to supervene on the physical. And if knowledge fails to supervene on the physical, then it can be caused by the mental without violating the Exclusion Principle. Thus, if knowledge is conscious, on this view, the mental can be causally relevant to our knowledge of it.

It is not obvious that this dual-explanandum theory will solve all the problems associated with Epiphenomenalism. Knowledge is not always conscious. Procedural knowledge, or know-how, seems to be largely unconscious. Perhaps even some propositional knowledge is unconscious, if we can have unconscious beliefs. Moreover, additional problems beset Epiphenomenalism that are not addressed by this solution. Specifically, Epiphenomenalism is incompatible with our ability to refer to our own mental states, but our ability to refer is not itself a conscious state (although some acts of reference may be conscious). So, a dual-explanandum view, based roughly on Jackson's idea that the mind can be causally relevant at least to first-person knowledge of it, only avoids some of problems with Epiphenomenalism. This dual-explanandum theory avoids problems of conscious knowledge of our own minds but not problems of reference to our own minds.

The third problem is that mental causation is only a nomic relation, so this reason for rejecting supervenience does not provide logical space for mental causation. Conceivability of absent or inverted qualia cases provide reason for rejecting supervenience only for logically necessary connections between the mental and the physical. The conceivability objection does not provide reason to reject claims of nomic necessity. It may be logically possible that qualia differ without a physical difference, but it is not, at least on this view, nomically possible that qualia differ without a physical difference. Since causation is only a nomic relation, this theory provides no room for the mental to be causally relevant since in all nomically possible worlds, the mental is still determined by the physical. The mental still supervenes strongly on the physical, on this view, when the strength of the modal operators is only nomic not logical necessity. Thus, this conceivability objection provides no solution to the problem of mental causation.

The fourth problem is that this solution does not give the kind of mental causation that we want. If logical space is made for mental causation by a failure of conscious states to supervene on the physical, then this theory only allows for the mental to cause conscious mental events. It does not allow for the mental to cause behavior or other nonconscious events. Thus, on this view, since my act of getting a beer is not itself a conscious event, my desire for a beer does not cause me to get a beer. My desire for a beer can only cause other mental events, like my knowledge or conscious belief that I desire a beer. So, this solution allows for only some mental causation and not all the mental causation we want.

The fifth problem is that this view seems to involve a mysterious parallelism between the mental and physical and, perhaps, panpsychism as well. If consciousness is

not hooked up to the physical in some way, then it is an extraordinary accident that consciousness correlates as it does with brain states. Moreover, the view suggests panpsychism. If the mental can be causally relevant because there is logical space for conscious states that are undetermined, then this theory suggests that any causes for conscious events are mental causes.

However, not all qualia have mental causes. Some conscious events, assumed for the sake of argument not to be physical events, say Tim's pain, are caused by physical events, events with no mental properties at all. For example, a rock falling on Tim's head causes him pain, which, on this view, does not supervene on the physical state that underlies it. Tim's pain is not determined, therefore, by the rock falling on his head. But if this is correct, either Tim's pain is accidentally correlated with the rock falling on his head or Tim's pain is caused by some nonphysical event correlated with the rock's falling. Neither of these possibilities is at all plausible. The first possibility suggests a mysterious parallelism between the mental and the physical. The second possibility suggests panpsychism. Thus, it must be that the rock's falling on Tim's head really causes his pain. Physical causation or determination of some conscious events allows for the world to hook up to conscious events to avoid panpsychism. Physical determination of some conscious events avoids the problem of parallelism for some but not all conscious events. But the idea that the physical determines some conscious events contradicts this theory's solution to the problem of mental causation. That solution depended on a failure of supervenience or determination of consciousness by the physical.

Moreover, if Tim's pain can be caused by a physical event, we have no reason to think that other mental events require mental causes rather than physical causes. If conscious mental events can be caused by physical events, even though it is logically possible that the conscious event is not determined by any physical event, there is no reason to think any conscious mental event needs a mental cause. For every conscious event, like Tim's pain, we are faced with a similar dilemma in deciding whether the event has a physical cause or correlate. It seems from neuroscientific evidence that every conscious event has some physical correlate. We must conclude either that each physical correlate is sufficient for the conscious event or that only some physical correlates are sufficient for conscious events. If we take the former horn of the dilemma, we no longer have a solution to the problem of mental causation. If the physical correlate is sufficient, then the conscious state is determined, and we no longer have logical space for it to have a mental cause. If we take the latter horn of the dilemma, then we need some theory for why some conscious events are determined by physical correlates and others are not. It is hard to see how one would begin formulating such a theory. Thus, this solution does not provide a reason to think the putative mental causes really are causally relevant.

So, supporting a dual-explanandum view for consciousness with the claim that qualia do not supervene on their physical realization will not provide a solution to the problem of mental causation. Even if the contentious claim that conceivability entails logical possibility were true, several apparently insurmountable problems would remain. These arguments seem obvious enough that this suggested solution should have been a nonstarter, and these arguments against it belabored the obvious. Nonetheless, the idea of a limited Epiphenomenalism, or a dual-explanandum view, that avoided some of the

problems with a full-blooded Epiphenomenalism seemed tempting. The only way to make use of this idea without rejecting the Exclusion Principle is to appeal to a failure of supervenience. These reasons for failure of supervenience do not provide a solution one can live with. So, we must look to other reasons for a failure of supervenience to provide a dual-explanandum response to the problem of mental causation.

7.4 Quantum Holism

The second reason one might have for doubting that the mental supervenes on the physical comes from quantum mechanics. Certain quantum systems have determinate properties although parts of those systems lack locally determinate properties⁸. For example, two particles emitted in opposite directions would individually lack determinate momenta, but, because of conservation of momentum laws, the two-particle system has a determinate net momentum of zero. Some quantum systems, thus, have holistic properties; whole systems have determinate properties while their parts lack locally determinate properties.

One might believe that the mind is itself a quantum system, as Eccles (1994), Popper and Eccles (1977), and Penrose (1989) apparently do. This response is implausible because it would appear to involve mysterious nonphysical mental forces acting directly on other mental events. The possibility that mental entities directly affect other mental entities without any physical implementing mechanism is quite mysterious and appears to involve direct causation from one ghostly entity to another. Moreover, since some quantum effects are, apparently, irreducibly random, this view would have

⁸ The caveat that the individual particles lack locally determinate properties is important. The individual particles may have determinate properties relative to a measurement, as claimed by Bohr's Copenhagen Interpretation, but the measurement apparatus is not a local property of the particle. Supervenience, as we have interpreted it, requires local physical properties as a supervenience base.

mental events occurring randomly. Since our thoughts and actions certainly do not appear to be random, the mind does not seem to be a quantum system. Thus, appeals to direct quantum effects of the mind are implausible.

A more plausible interpretation of this quantum holism is that the failure of fundamental features of the world to supervene on their parts should shake our confidence in the reductionism philosophers often read into science. Perhaps the world is rife with failures of supervenience, the claim goes, so reasonable naturalists should not conclude that every layer of the putative hierarchical organization of the world supervenes on the one below; nor should philosophers conclude that all special science properties are determined by their realization bases.

A dual-explanandum theorist could take advantage of this failure of supervenience by appealing to a hierarchical series of causes each independent of the one below it. I will call this view a "holist dual-explanandum" view. Thus, the mental could cause the mental because no physical (or chemical or biological) cause determined the mental. Because Fred's action of getting a beer does not supervene on any locally determinate properties of the parts of his action, it becomes possible that his desire for a beer can cause his action. The layered structure of the world, thus, could involve a layered set of causal relations, each independent of those below.

There are two problems with this theory. First, the evidence for this theory of the structure of the world is weak. Adducing any remotely conclusive evidence for or against this interpretation of the evidence from quantum holism goes far beyond the scope of this work, and is, perhaps, not yet empirically available. However, there is some reason to think this anti-reductionist worry is not well-supported by the existing

evidence. These failures of local determinism in quantum mechanics have seemed so implausible as to be nearly incomprehensible. Many respectable physicists, including most notably Einstein, were simply unable to accept the failure of local determinism. The very revolutionary nature of the evidence from quantum mechanics suggests that such failures of local determinism are not common in the sciences. Essentially, quantum mechanics raised such problems for a realist view of the world precisely because such radical failures of supervenience are extremely rare and were previously unseen.

Moreover, extensive evidence from neuroscience involves perfectly determinate properties of the brain--activation of certain neural centers, release of certain neurotransmitters, damage to specific areas of the brain--that correlate with determinate mental properties. Thus, at least in the case that most interests us, the case of mental causation, the evidence suggests that brain properties are determinate and, therefore, the evidence suggests that quantum holism does not provide reason to reject supervenience of the mental on the brain⁹.

The second problem with appeal to quantum holism is that, even if this theory were correct, it would not provide a good theory of mental causation. First, we need to be clear on how this theory might allow for mental causation. On this theory, the brain, or parts of it, would not have determinate properties at the same time that the mind had determinate properties. For example, I could have a pain but my brain, or parts of it, may have no determinate property corresponding to that pain. In addition, my belief that I am in pain, or perhaps my pain behavior, would also be determinate while the underlying brain states or physical movements would not be determinate. Thus, on this view, my

pain could cause my belief that I am in pain without any brain states determining the belief because, in fact, there are no determinate brain states to determine the belief. For quantum holism to allow for mental causation, mental items would have to be determinate without determinate brain states, and those mental items could cause or quause other mental items or behavior for which there is no determinate underlying items. Evidence from quantum holism seems to allow for a dual-explanandum view of mental causation.

This holist dual-explanandum theory leads to the same worries (from section 7.3) about parallelism and panpsychism that the theory based on failure of consciousness to supervene on the physical did. The parallelism objection is that it appears that the mental correlates mysteriously with the underlying physical or brain properties. The panpsychism objection is that if the only reason we can give for mental events to be causally relevant is that events occur that do not have physical determiners, then external causes of mental events appear to be themselves mental. This possibility suggests that mental events properties occur external to anyone. These problems suggest that a holist dual-explanandum view will not allow for mental causation.

The parallelism objection to the holist dual-explanandum view is that the correlation between mental and physical items appears completely mysterious. If there is no determination from the physical to the mental or mental to physical, then it is mysterious how our mental lives correspond so closely to the activities of our brains. On this view, no determinate physical events correlate with the mental or macroscopic event, so there is nothing physical for our mental lives to correspond to. The holist dual-

⁹ It might be argued that this evidence is only evidence for observed properties of the brain. Once the brain is observed, even on the quantum theorist's view, the brain has determinate properties. But this brain

explanandum view appears to avoid this problem, but the attempt to avoid it raises other problems.

This response raises problems of its own. First, mental events or properties really do correspond to brain events or properties. The evidence from neuroscience, for example from PET (Positron Emission Tomography) scans that allow for observation of activity in the brain while the subject is conscious, shows a correspondence between the mental and brain activity. The holist dual-explanandum theorist would have to respond that PET scans provide examples only of *observed* brain activity, not simply brain activity. The evidence from quantum mechanics is that physical properties are indeterminate only until observed. So, the case of correspondence between the mental and the brain during a PET scan cannot count against the view because the view only claims that unobserved brain events have no determinate properties.

However, this response is inadequate because it implies that mental events could not be causes when the brain events are being observed. For example, the brain event underlying my belief that I am in pain is completely determinate when my brain is observed, so my belief, in those circumstances, is completely determined by the brain. So, assuming the Exclusion Principle, my pain cannot also determine my belief that I am in pain. So, on this holist dual-explanandum view, the mental would become epiphenomenal whenever the brain is observed. But the problems with Epiphenomenalism are quite general and apply even when one is undergoing a PET scan. Thus, the holist dual-explanandum view only avoids the parallelism objection by supposing that mental events are undetermined only when their underlying brain events

research has not discovered any of the indeterminacy effects that make quantum mechanics so mysterious.

are unobserved, but that implies that Epiphenomenalism is true when the brain states are observed.

The problem of panpsychism is that the holist dual-explanandum view appears to countenance mental entities external to brains. Many mental (or special science) events are caused by physical (or lower level) events. To repeat the example from 7.3, a physical event, a rock falling on Tim's head, may cause a mental event, Tim's pain. To think that Tim's pain is not caused by the rock we must either think that Tim's pain and the rock falling on his head are accidentally correlated in some mysterious way or we must think that Tim's pain was caused by some mental event correlated with the rock's falling. We have already discussed the problem of parallelism. The other solution is to suggest that the rock has some mental property or that its falling is somehow a mental event. This view is panpsychist; it countenances mental properties in rocks and other physical objects that we do not think of as mental. Since panpsychism is unacceptable, some mental events must have physical, nonmental, causes or determiners. But it is mysterious why some mental and behavioral effects have mental causes and others do not, on this view. If the mental is not required to cause mental events, like Tim's pain, then it is hard to see why the mental is necessary as a cause for any mental event. One could attempt to give some theory explaining why a mental cause is necessary for some event but not others, but it is hard to imagine how such a theory would go. So, even if this somewhat dubious inference from quantum holism to a general claim about the structure of the world, the holist dual-explanandum view, were correct, it would not provide a plausible way for mental causation to occur. We should, then, reject theories of mental causation based on evidence from quantum holism.

7.5 Individuation by Context

The third, and best accepted, type of evidence for a failure of supervenience involves individuation of the mental or special sciences at least in part by context. I will not yet describe a particular dual-explanandum theory which might take advantage of this failure because so many dual-explanandum theories have been proposed. Instead I will first describe the evidence that, because of the context-dependence of at least some content or mental properties or events, the mental does not supervene on the physical.

Evidence of the role of context in individuating content or other properties only shows that weak and strong supervenience fail for these properties. Context can be fixed by fixing all the physical properties and their distribution for a world. So the following evidence constitutes a reason to reject weak and strong supervenience but no reason to reject global supervenience.

The first main area of evidence involves mental content. Mental content properties do not supervene on neurophysiological (physical) properties. For example, Putnam (1975) has argued that meanings 'ain't in the head.' The meaning of my thought that 'water is wet' is not individuated by what is in my head, either physiologically or psychologically, but by my causal relations to the physical substance H₂O. A person with a physiologically and psychologically indiscernible brain on Twin Earth in which the lakes are filled with some other chemical, say XYZ, would have thoughts about XYZ and not H₂O. So, the meaning of my word, 'water' and the meaning of whatever token occurs in my brain, say #water#, is individuated not by local physical properties of mine, for example my neurophysiology, but by the causal context in which my word or thought

occurs. Putnam's examples are widely thought to show that meaning cannot be individuated by local physical properties¹⁰.

Burge (1979) argues that mental content, such as the belief that I have arthritis, is individuated (at least in part) by the society of which I am a part. Meanings of mental tokens are fixed by the society, so whether my thought 'I have arthritis,' is actually the thought that I have arthritis or whether it is the thought that I have some different ailment, which we can call *tharthritis*, is individuated by facts about the society of which I am a part. Putnam's and Burge's arguments show that mental content or meaning of mental tokens does not supervene on local physical properties.

Davidson (1986) has argued that intentional states, such as beliefs and desires, are individuated by causal relations to objects in the world¹¹. For example, a molecule-for-molecule duplicate of me that comes into existence in the middle of a swamp by a cosmic accident could not have the desire to visit Vienna because it has no causal relation to Vienna. I may have the desire to visit Vienna because I am causally related to Vienna, but my double is not so related to Vienna. Davidson adds the claim that mental content individuates mental states like beliefs, desires and thoughts. Since the content of these states does not supervene on local physical states, the states themselves do not supervene

¹⁰ The original purpose of the example was to show the anti-Fregean point that there can be no one thing that is both the meaning of a term and the thing we grasp when we understand the term. The example of water and twin-water makes this point. The stronger conclusion above is widely thought to follow as well.

¹¹ This claim appears to contradict Davidson's (1970) claim that the mental weakly supervenes on the physical. I do not know whether he sees these claims as contradictory. According to weak supervenience, as long as beings in the same world share the same physical properties, they have the same mental properties, but there is no necessity that beings in other possible worlds that share physical properties with beings in the actual world should have the same mental properties as beings in the actual world. As long as the swamp-thing is a merely possible individual, then there is no outright contradiction with weak supervenience. This move seems implausible, however, and it seems unlikely that Davidson would adopt it. Moreover, supervenience fits uneasily with Davidson's radical interpretation as individuating of mental states because it is easily possible on that view for there to be a change in interpretation without a physical change.

on the physical. One could have different intentional, and hence mental, states without having different local physical states. Davidson, thus, is committed to a claim that goes beyond those of Putnam and Burge. He must be committed to the claim that the mental does not supervene on the physical because the content does not supervene on the physical. Whatever Davidson's additional commitments, three independent, and apparently convincing, arguments make the case that mental content does not supervene on local physical properties.

In addition, at least some special science properties do not supervene on their realization bases. Fodor (1974) argues that special sciences and their laws are autonomous of the implementing mechanism. Fodor argues that laws of special sciences like economics require appeals to properties that are multiply realizable in heterogeneously disjunct physical states. No single type of physical property is likely to pick out the properties necessary for laws of economics. For example, laws of economics appeal to money, and money is not identical with any type of physical item. Wampum, credit cards, and green pieces of paper have no physical properties in common that might be relevant to their being money. So, money is not reducible (i.e. identical) to any physical type.

A further conclusion one might draw is that money is individuated contextually. Given Fodor's examples, it appears that special sciences may appeal to context-dependent properties. Not only do realizations of money not have any physical property in common, no physical property is sufficient for something to be money. Money is individuated by its role in society. Green pieces of paper indistinguishable from

Davidson's thought experiment here may be the weakest of the three arguments. Some philosophers simply do not share Davidson's intuition in this case.

American dollar bills would not be money in other contexts. So, money does not supervene on any local physical properties. Two objects could be indiscernible in all local physical details but one could be money because it occurs in one context, and another might not be money because it occurs in a different context. So, the special sciences often appeal to real properties that do not supervene locally on physical properties. So, Fodor's (1974) argument for the irreducibility and autonomy of the special sciences provides a reason to reject the supervenience of the special science properties, including mental properties since psychology is a special science, on their physical realizations.

A minimal conclusion of these arguments is that at least content properties and possibly some mental properties do not supervene on local physical properties. Instead, these special science or content properties are individuated at least in part by the context in which they occur. For my thought to be a water-thought requires that I be in the correct context. Whether the context is a causal-historical or social context does not matter for the discussion here. Whatever theory, if any, turns out to be correct, content properties are not individuated locally; they depend on some sort of context. My further conclusion from Fodor's argument shows that it is possible that mental properties, insofar as they are fixed by their context, fail to supervene on local physical properties.

As I noted in the first chapter of this work, there are often thought to be three distinct problems of mental causation for Nonreductive Physicalism. These problems are, first, the problem generated by Davidson's Anomalous Monism, second, the problem generated by the extrinsic nature of mental content, and third, the problem of mental causation generated by the Exclusion Argument. The second problem is the one mainly

concerned with mental content as individuated widely or extrinsically. Thus, the extrinsic nature of mental content is often thought to raise the problem of mental causation rather than solve it.

One could make the latter two arguments part of a single argument in dilemma form (as Kim [1998] does). If content is intrinsic, then it supervenes on the physical and thus the Exclusion Argument shows that it is causally irrelevant. If content is extrinsic, then it does not supervene on the physical, and because all causation is local, extrinsic content is causally irrelevant. Either way, content is causally irrelevant. Thus, although it is tempting to separate these two problems here, it is important to the problem of mental causation to show that a solution to one problem does not simply land one in the middle of the other problem.

7.6 Some Inadequate Solutions

As I have argued, for a dual-explanandum theory to be correct, it must include a rejection either of the Supervenience thesis or the Exclusion Principle. Any of these reasons for the failure of supervenience would give reason to reject the Exclusion Argument as unsound, but they do not necessarily provide an explanation for how the mental might be causally relevant. I will briefly consider three suggested solutions that seek to bypass the Exclusion Argument. Then I will consider at greater length one solution that tries to confront the Exclusion Argument head on.

7.6.1 Burge's Solution

Merely observing that mental content does not supervene on the physical does not in itself constitute a solution to the problem of mental causation. For example, Burge (1993), correctly on my view, accepts that mental content does not supervene on the physical but

does not see this failure as impacting the problem of mental causation. Burge relies on explanatory practice and our ordinary notion of causal powers to allow for mental causation. He argues that common explanatory practice picks out some regularities as causal, and since this explanatory practice assumes mental-to-mental causal relevance or mental-to-physical causal relevance, mental properties are causally relevant. Burge realizes that relying on regularities alone fails to distinguish epiphenomena from causally relevant properties, so requires instead that common explanatory practice be our guide in picking out causal relevance.

However, Burge's solution leaves something to be desired. First, his assumption that common explanatory practice assumes causal relevance may simply not be true, and is certainly not justified without additional argument. Second, he does not provide an explanation to the philosophically more important question of *how* mental causation may occur. Third, common practice is defeasible, and the Exclusion Argument may provide reason to defeat it.

Kim (1995) points out the first two problems with Burge's appeal to common practice. First, it has not always been accepted that rationalizing explanations are causal explanations, so common practice does not obviously assume causal relevance. For a significant part of this century philosophers¹² thought that rationalizing explanations were not a variety of causal explanation at all. The reasons cannot be explored here in detail. The main arguments involved the logical and normative connections between an action and an intention. For example, Malcolm argues that intentions are logically connected with actions in a way that causes cannot be. On a causal view, he argues, I could have a desire to drink a beer without drinking the beer even if all conditions are ideal. However, if all the conditions

are ideal, and I do not drink the beer, Malcolm claims, I do not have the desire to drink the beer. It is part of the concept of desire that I cannot have the desire without acting upon it, *ceteris paribus*. Causal relations do not have this feature. It is not part of the concept of a cause that it bring about a certain effect even *ceteris paribus*. Thus, my desire to drink a beer, on this view, does not cause me to drink the beer.

Moreover, normative considerations apply to reason explanations but cannot apply to causal explanations. Reasons cannot cause actions because of the normative character of reason explanations. Take, again, the example of my drinking a beer. My reason for drinking a beer may be good or bad, justified or unjustified. But a cause can be neither good nor bad, neither justified nor unjustified. These arguments may be ultimately unconvincing. I do not have the space to consider the arguments as carefully as required to determine their soundness. However, they are sufficient to show that one cannot simply assume that the common practice of intentional and reason explanations are causal explanations. Common practice does not in any obvious way assume mental explanations are causal.

Second, as Kim also points out, even if explanatory practice assumes that the mind is causally relevant, this fact does not explain how it is possible for the mind to be causally relevant. The problem of mental causation is not that we do not think the mind is causally relevant but that we do not have a metaphysical picture of the mind and the world that allows for the mind to be causally relevant. The question is not so much whether the mind is causally relevant, but rather how it is possible for the mind to be causally relevant. And this question Burge's appeal to common practice does not answer.

A third problem for Burge's argument is that our common practice may be mistaken. In this case, we may mistakenly attribute causal relevance to mental properties. Scientific

¹² For instance, Anscombe (1963) Malcolm (1968), and Wittgenstein (1958)

considerations have often overcome common practice. Perhaps the case of mental causation is another case in which scientific considerations, suitably informed by philosophy, should overcome our common practice. These three reasons are sufficient to reject Burge's solution. Appeals to explanatory practice alone cannot be sufficient to solve the problem of mental causation¹³. Thus, some substantive theory of mental causation that takes into account the Exclusion Argument is necessary to solve this problem of mental causation.

Although rejection of the Supervenience thesis or the Exclusion Principle is necessary for a dual-explanandum view, among those philosophers who reject supervenience, few have given theories, let alone dual-explanandum theories, of how the mental could be causally relevant. Lynne Rudder Baker (1993) and Jerry Fodor (1989) both give dual-explanandum theories that are compatible with a rejection of only the Supervenience thesis. Baker (1993) rejects any metaphysical principle, preferring to rely on common practice. Fodor (1989) does not specify which premises of the Exclusion Argument he finds objectionable. Both theories, however, rely on appeals to causal powers as determined by laws, and because these causal powers and laws, given the arguments above, may not supervene on their physical realizations, these theories may benefit from a rejection of the Supervenience thesis without also requiring rejection of the Exclusion Principle. Therefore, these theories are properly categorized as dual-explanandum theories in which special science entities cause other special science entities independently of their implementation.

Baker and Fodor need not deny the Exclusion Principle. The Exclusion Principle only applies to events or properties whose occurrence is determined by more than one

¹³ Common explanatory practice may, however, be part of an argument for a particular solution. For example, it may provide a reason to accept a dual-explanandum view.

complete set of causes or causally relevant properties, and if content, mental or special science properties do not supervene on the physical, then the physical does not determine the mental. The point here is not to interpret Fodor's or Baker's actual view but to show that their dual-explanandum theories require only rejection of the supervenience thesis.

7.6.2 Baker's Solution

Baker (1993) suggests that we take laws to give causally relevant properties. Laws of nature, on this view, advert to a property of a cause which gives rise to some further properties or events. The problem with this theory is that not all laws are causal laws. As Fodor (1989) points out, the ideal gas law is clearly not a causal law. The ideal gas law says that for an ideal gas (one without interactions among its particles) the pressure, volume and temperature of the gas are related in a law-like way, but the pressure does not cause the temperature to be such and such; nor does the temperature cause the pressure to be so and so. Instead, the law represents a macroscopic generalization from the facts of statistical mechanics of the particles in the gas. So, Baker's appeal to laws will not solve the problem of mental causation as presented by the Exclusion Argument.

7.6.3 Fodor's Solution

Fodor (1989) notes that not all laws are causal and so suggests instead *causal* laws as grounding causally relevant properties. The properties adverted to in causal laws are causally relevant, on Fodor's view. And Fodor is confident that some notion of causal law can be given independently of a notion of causal powers or causal relevance. If Fodor is correct, then mental causation can be given a grounding independent of its physical realization or implementing mechanism.

It might be objected that Fodor's proposal gets the relationship reversed. Perhaps entities fit into causal laws because they have causal powers, not *vice versa*. In other words, one might think that the causal powers are the fundamental existents, and the causal laws are parasitic on them¹⁴. Determining whether the properties or the laws are more fundamental is a difficult question, the pursuit of which is outside the scope of this work. Thus, I will criticize Fodor on another front.

Even if causal powers depend on causal laws, it is far from clear that a reductive theory of causal laws is possible. It is hard to see how any logical or modal relation could capture the metaphysical difference between a genuine causal power and a nomically correlated common effect. Epiphenomena could be nomically correlated with actual causal powers in such a way that no formal distinction between them could be found. For example, it may be a law of nature that all renates are chordates, but the two properties, having hearts and having kidneys, are clearly distinct. The property of being a renate may cause certain physiological effects whereas the property of being a chordate does not have these effects. But no nomic correlation could distinguish this causal relationship from a nomically indiscernible relationship between the property of being a chordate and those effects. No modal relation or asymmetry relation is likely to distinguish one property as causally relevant and the other as causally irrelevant.

One might respond on behalf of Fodor's claim, assuming that special science properties do not supervene on the physical, that no such nomic correlations could hold between physical and special science properties. This response, however, misses the point of the objection. The objection is that Fodor could give no characterization of a *causal* law without assuming the *causality* of the law itself. Whatever formal, modal or

¹⁴ Barbara Hannan (1994) argues in this vein.

logical relation Fodor thinks might capture the causal power of the law is simply not likely to determine that laws with identical formal features involve properties that are causally relevant rather than epiphenomenal.

This brief review of some recent attempts to solve the problem of mental causation suggests that attempts to bypass the Exclusion Argument are not likely to succeed. Any attempt to solve the problem of mental causation, the exclusion problem, should deal directly with the soundness of the Exclusion Argument.

Dual-explanandum theories must address the Exclusion Argument. Those solutions that work by rejecting supervenience will have to make use of the context-dependence of mental content or special science properties. Rejecting supervenience because of conceivability of absent or inverted qualia and rejecting supervenience because of quantum holism both lead to problems with connections between the physical and the mental. Neither of these theories explains how such connections are possible. Dual-explanandum theories that appeal to a failure of supervenience because of context-dependence, on the other hand, allow for connections between the mental and the physical. A physical item within the proper context determines the mental or content item, on this view. Thus, there is no mystery about how the mental and physical are related; they do not require a mysterious parallelism. I will consider one further attempt to solve the problem of mental causation by explicit appeal to failure of supervenience.

Recall that there is more than one problem of mental causation. One problem is that presented by the Exclusion Argument. Another problem is that presented by Davidson's Anomalous Monism. Still a third problem is that presented by the failure of mental content to supervene on the physical. This third problem is raised against those

philosophers who believe that mental content is not individuated locally but who believe all causation is local. These two assumptions appear to show that mental content, because it fails to supervene on the physical, cannot be causally relevant. Several philosophers have tried to solve this last problem. A sketch of a general solution to the problem of mental causation for extrinsic properties can show how such a solution might be applied to the problem of the Exclusion Argument. However, I will argue that such a solution cannot solve the problem of mental causation for the Exclusion Argument without rejecting some other premise of the argument.

7.7 Wide Content and Wide Behavior

The dual-explanandum view I will discuss here instead of worrying about the failure of mental content and wide behavior to supervene on the physical makes use of this failure¹⁵. The failure of wide behavior to supervene leaves logical space for mental content to determine wide behavior. Purely physical states of the brain cause narrow behavior, but wide content causes wide behavior.

On this theory the context of a mental token is causally relevant to wide behavior but not to narrow behavior. The distinction between narrow behavior and wide behavior is the following. Narrow behavior is simple bodily movements, whereas wide behavior is behavior that consists of more than simple bodily movements¹⁶. Consider, for example, my wide behavior of making my bed. I can go through the narrow behavior involved in making my bed without actually making my bed. For example, I could make someone else's bed, or perhaps I could mime making a bed. My wide behavior of

¹⁵ No solution in the literature closely resembles the one I suggest here. Heil and Mele (1991) and Dretske (1988) provide dual-explanandum theories that are directed at the problem of the causal relevance of wide, extrinsically individuated, content, but these views do not directly address the problem raised by the Exclusion Argument.

making my bed can only occur when I actually make *my* bed. On this theory, without my behavior occurring in the proper context (in this case, intuitively, a causal-historical relation to my bed), the wide behavior would be different. If I weren't causally related to my bed, then I would not be making it, but doing something else. Thus, the context, possibly including the mediate causal history, that individuates mental content is causally relevant to wide behavior.

Noting the causal relevance of context does not solve the problem of mental causation for extrinsic content, however. Context may individuate content, but, unless content is type or token identical with that context, then content itself has not been shown to be causally relevant. Dretske (1995), for example, endorses the type identity of mental content with the context that individuates it. He writes,

Content, after all, is extrinsic. If it reduces to anything, it reduces to some set of physical relations existing between what is inside the head and what is outside. This is a *type-type* identity theory, but not the classic (Smart-Feigl) identity between a type of mental state (e.g. the belief that F) and a type of brain state, but, rather, an identity between intentional content and the set of external relations that give brain states their representational character. This is reductionism, yes, but (if you please) *wide* reductionism. (1995, 147-48)

Accepting the causal relevance of context, alone, tells us nothing about the causal relevance of the mental content. Given the existence of causally relevant context, the context, and the physical states of the brain, may be sufficient for any effects that occur. Thus, assuming the Exclusion Principle, mental content could be an epiphenomenal effect of the genuinely relevant context. To avoid this conclusion, without rejecting the Exclusion Principle one must assume that mental content is identical to the context itself.

To put the problem another way. If we assume that the only reason supervenience of the mental on the physical fails is because of the context-dependence of

¹⁶ This distinction comes from Heil and Mele (1991).

behavior and the mental, then it is possible for the Exclusion Argument to exclude the mental yet again. Fixing the physical context of the mental event must be sufficient to fix the content of the mental event. If the mental globally supervenes on the physical, then it appears possible to exclude the mental properties of the world as a whole. For example, the physical context of the my desire for a beer determines that the desire be for a beer, but the content of that desire need not be causally relevant (or quausally relevant) to my drinking *being a beer-drinking* unless the content is identical to that complex set of relational physical properties. The complex set of physical relational properties is sufficient to bring about the beer-drinking, and given some physicalist theory of content, they are also sufficient to making the drinking a *beer-drinking*. Thus, the wide behavior can be determined by the physical and the physical context, so, applying the Exclusion Principle, since there is a complete physical cause, in this case a complex set of relational physical properties, the mental event cannot be causally relevant.

The only solution, then, is to appeal to an identity between mental content and the context that individuates it. In this way, we can preserve the mental token's causal relevance since it is identical to something that is admittedly causally relevant. Before considering how an identity theory might work, we need to consider what items are thought to be identical. The identity theory can be either an identity of content types or content tokens. The type identity theory would identify content types with types of complex physical relations. For example, the content of my desire for beer is that it is a desire for beer. The physical relations underlying this content are likely to involve some causal or nomic connections between me and beer (and possibly my linguistic community). The type identity theory for content is that this connection, whatever its

nature, constitutes a type which is identical to the content type *beer*. An identity theory of content tokens would be the view that this instance of my desire for beer has as content some particular causal or nomic relation between me and beer.

On this theory, then, content is a property of an event, and that content is identical to a complex physical relation. The mental tokens fall under the content type. Say, my desire for a beer causes me to get up and go to the fridge. The content of that desire is that it is a desire for a beer. That content is identical to some complex physical relation between persons and beers. My desire falls under that content type because it falls under that complex physical type. And the content of my desire is causally relevant because the complex physical property it is identical to is causally relevant.

A type identity theory, like the one Dretske (1995) espouses, solves the problem of mental causation for content¹⁷. If the mental content type is identical with the causally relevant context that individuates it, then the mental content type can be causally relevant. The mental type, since it is identical to a physical type, can cause physical events. Moreover, since the mental type and the complex physical type are identical, physical events occur in virtue of that single type. There is no question of whether events occur in virtue of the physical type or the content type since they are one and the same. If mental content is type identical with a physical type, then the Exclusion Argument cannot have even the minimal property dualism it requires.

¹⁷ It is not clear how Dretske intends this identity theory to help in his (1988) solution. For Dretske (1988) behavior is a causing of a physical movement. My behavior is something I do rather than something that happens to me. Content, on his view, is the structuring cause, or the background conditions that make an event a behavior rather than a simple movement. The content of my desire, as a beer-desire, is what makes my movements beer-getting behavior. However, the identity claim he espouses in his (1995) seems to accord to content, because of its identity to physical relations, a role as an ordinary, or in his terms, *triggering*, cause of a movement. Thus, Dretske's solution in his (1995) does not appear compatible with his general theory of mental causation in his (1988). Nonetheless, Dretske's response seems to help with the problem presented here.

The type identity claim is familiar from the Exclusion Argument. The difference here is that the identity claim is not the identity of a mental event with a brain event; instead the identity is between mental content and a complex physical relation. We have no clear idea what this complex physical relation might be because we have no clear idea of what physical relation individuates mental content. Nonetheless, this theory claims, whatever complex physical relation individuates content is type identical to the content itself.

However, it is unlikely that these complex physical types are identical with mental content types. One can raise several objections to this view. First, what individuates something is generally not identical to it. Second, it is likely that content is realized in heterogeneously disjunct sets of physical relations. Third, this solution works only for mental content and not for qualitative conscious states.

The first objection is that what individuates a thing is generally not identical with it. For example, the individuating conditions for a species of animal may involve the animal having certain physical structures, e.g. a jaw muscle that attaches to a certain place on the jawbone, and have certain reproductive habits, e.g. individuals who interbreed are considered part of the same species. But these individuating conditions themselves are clearly not identical to the species. Assuming that context does individuate content, it remains to be argued that context is identical to that content. Perhaps the only argument for this claim is that it solves the problem of mental causation.

The second problem with the type identity claim is simply that the identity does not seem to hold because of multiple realizability. Mental content, say that my desire is for a beer, is realizable in a wildly disjunctive set of relations between brain tokens and

beers. Every possible case in which someone desired a beer, on a type identity theory, would have to instantiate this complex physical relation, but it seems unlikely that this relation could obtain for every brain token #beer# and the beer itself. It is likely that #beer# tokens arise in a heterogeneously disjunct set of ways with no physical type in common. For example, my thoughts about beer might arise from first hand experience with Budweiser ("The King of Beers"), whereas an extremely devout Mormon may never have tasted or even seen beer first-hand but only read about it in church publications. It is unlikely that any single type of physical relation could cover these two cases and all the myriad other ways one might come to have thoughts about beer. No type of physical relation or connection seems to cover all the possible ways one's mental tokens could come to represent an object.

One difficulty with raising this objection is that no one has a finished theory of meaning or mental content that is free from counterexample. No theory has given uncontroversial necessary and sufficient conditions for individuation of content. Until that has been done, it is impossible to attempt to give a counterexample to the identity claim. However, the fact that no theory of content is without significant problems or counterexamples shows, I think, that no identity theory for content is possible since some set of individuation conditions would be necessary for an identity claim.

The other variety of identity theory that remains possible is an identity of mental content tokens with complex physical relation tokens. On this view, one instance of mental content is identical with some particular physical relation. Thus, my desire now for a beer is identical to a complex relation token between me and a particular beer (and a particular context). This view avoids the multiple realizability objection because it does

not postulate some single physical type common to all beer-thoughts. On this view, a mental content token, because it is a physical relation token, can cause a physical effect. So, postulating a token identity between content and a complex physical relation appears to solve the problem of mental causation for the Exclusion Argument¹⁸.

Certain questions remain, however. Specifically, we still need to know whether it is in virtue of the content token's falling under a content type or falling under some physical, relational type that the effect occurs. A token identity theory avoids concluding Token-Epiphenomenalism, but it does not necessarily avoid concluding Type-Epiphenomenalism. For example, when one wonders about the causal relevance of my pain token, one can ask whether, when I hit my thumb with a hammer, the pain event caused my screaming or whether my screaming occurred in virtue of the event being a pain. Certainly a token identity theory does away with worries about whether the pain can cause my screaming (where, one supposes, the screaming is also at least token identical to a physical event). However, this response does not answer the question of whether my pain, *qua* pain, causes my screaming, *qua* screaming. Similarly, one can ask whether desire, *qua* beer-desire, is relevant to my physical movements *qua* beer-drinking.

The only way this theory can work, as far as I can tell, is if the content (or mental) properties and the physical properties cross-classify the same event tokens. Thus, my bodily movement is both a beer-drinking and an arm and throat movement. There is only one event here with two distinct properties. The event is caused by a physical event token which also has two distinct properties, one is a content property of the desire being a desire for beer, and the other is a complex physical relational property. One event

¹⁸ Taking this evidence to show that only a token-identity theory can be correct, and thus to show that constitution physicalism is false, would be premature, however, since I will find reasons to reject this

causes the other unproblematically since they are both physical events, but the event *qua* beer-drinking occurs in virtue of the desire *qua* beer-desire, and the event *qua* physical movement occurs in virtue of the cause *qua* complex physical relation.

The content property and the complex physical relational property are distinct types, which can be causally (or quausally) relevant, but are also identical tokens, and so can cause physical events. I would be happy to endorse this solution for content properties, and any other properties individuated by context, except for two problems. First, the content property of the effect is already fixed by the physical properties of that effect, and so there is no need for the content properties of the cause to be relevant. Second, even if this solution worked for content, it would provide only a partial solution to the problem of mental causation because it would not help with causes or properties that supervene on the physical.

The first problem with this solution is as follows. We noted that this solution to the problem of mental causation only denies weak and strong supervenience. Content properties are still thought to globally supervene on the physical. Context has a role in individuating content properties, but that context may be entirely physical context. The physical cause and the correct physical context, fix the content property of the effect. But if the content property of the effect is fixed, or determined, then, by the Generative/Determinative Exclusion Principle, that property cannot have another determiner. Thus, the content property of the effect cannot have the content property of the cause as a determiner. The content token can cause, but the content property of that token cannot be causally relevant to the content property of the effect. This is so because

solution.

the effect's content property is already determined by the physical properties of the cause and the context, the complex relational property or set of relational properties.

Thus, identity theories for content are inadequate solutions to the problem of mental causation. The type-identity theory is unlikely to be true because of multiple realizability and the token-identity theory does not solve the problem of mental causation raised by the Exclusion Argument. So, identity theories for content are unlikely to solve this problem of mental causation. And, because of this fact, theories that appeal to mental content's failure to supervene cannot solve the problem of mental causation raised by the Exclusion Argument.

Finally, even if this identity of content with a complex physical relation were possible, this solution would only solve the problem of mental causation for mental content. This theory provides no solution for the problem of mental causation for qualia or other mental states that are individuated narrowly. And, contrary to philosophers like Lycan (1996) and Tye (1995) and others, who argue for a representational theory of the mind, some mental properties, in particular qualia, are not individuated widely.

Qualia Externalism is, on my view, incorrect. Qualia Externalism is the view that qualia are individuated widely. My disagreement with these views is that they seem to rely on realism about the represented properties or objects. For example, on this view, my red quale represents a real feature of the world, redness. This view seems plausible enough if there really is such a feature as redness. However, the items in the world that are sensed as red are a heterogeneous lot, and worst of all, some of these features of the world are uncontroversially not red, as is the case in the case of certain visual illusions. Bidwell's ghost, for example, is an illusion in which one perceives a bluish-green color,

but the apparatus that presents the bluish-green image is a “half-black, half-white disk, with a slot through which a red lamp flashes.” (Hardin, 1990, 555) This example is by no means the only example in which one perceives a color that is objectively not in the real world, and the visual modality is not the only modality in which such illusions occur although it is the most widely-discussed in philosophy. It does no good to claim, as Lycan (1987 and 1996) does, that these qualia represent unactualized possibles because these illusion cases completely parallel cases of veridical perception except that the perceived colors do not exist. For example, in the case of Bidwell’s ghost, something in the world causes the quale just as an actual bluish-green object might cause a quale. But the apparatus in Bidwell’s ghost is in no way bluish-green. These cases are completely parallel, and yet Lycan’s analysis would treat them as distinct. So Lycan’s analysis is incorrect. Cases like these show that realism about color and other so-called secondary properties is unlikely to be true, and so the realism about these properties that qualia externalism relies on is unlikely to be true.

If I am right that qualia externalism is incorrect, the Exclusion Argument would create a problem of mental causation that could only be solved in a piecemeal manner. This solution based on a failure of supervenience would provide one solution for mental content and perhaps a different solution, if one is to be found, for other mental events or properties, like qualia, that are individuated narrowly.

7.8 Conclusion

The claims of this chapter are the following. Dual-explanandum theories fit with our explanatory practice and provide the only solution to the problem of mental causation that avoids postulating downwards causation. Dual-explanandum theories, however, to

solve the problem of mental causation must postulate either the failure of supervenience or the falsity of the Exclusion Principle. The dual-explanandum theories devised to take advantage of potential failures of supervenience either raised other problems or failed to solve the problem raised by the Exclusion Argument.

I considered three possible reasons one might have to reject the supervenience thesis, and discussed dual-explanandum theories that might take advantage of these failures. The first theory I rejected supposed that the conceivability (or apparent conceivability) of absent or inverted qualia suggested that qualia did not supervene on the physical. This theory raised several problems including that it involved a mysterious parallelism between the mental and the physical and involved a variety of panpsychism. The second theory I considered supposed that quantum holism might provide reason to reject supervenience at the macro-level, including rejecting supervenience of the mental on the physical. I rejected this reasoning because of the weakness of its empirical claims and because it had the same parallelist and panpsychist problems as the previous theory.

Finally, I argued that dual-explanandum views that rely on mental content's failure to supervene could not solve the Exclusion problem. These dual-explanandum theories must assume an identity between the context and the mental content. A type identity theory, I argued, was unlikely given the probability of realization of content in a heterogeneously disjunct set of physical relations. A token identity theory for content, on the other hand, appeared not to solve the problem of mental causation as raised by the Exclusion Argument. Neither identity theory solved the Exclusion problem. So, appeals to failure of supervenience for mental content do not provide a causal role for content.

Theories of mental causation that rely only on failures of supervenience do not solve the problem of mental causation as it is raised by the Exclusion Argument. Dual-explanandum theories, like those discussed in this chapter, require in addition rejection of another problematic premise, the Exclusion Principle. I will turn to a discussion of the Exclusion Principle in the next chapter.

It is important to note before discussing the Exclusion Principle that a general strategy of applying modus tollens to the Exclusion Argument cannot be maintained for the case of mental content or other properties that fail to supervene on the physical. If one admits that the Supervenience Thesis is false, one cannot then take the validity of the Exclusion Argument to show that some other premise of the argument must be false. Thus, I will not pursue such a strategy, but will instead pursue independent reasons for thinking the Exclusion Principle is either false or allows the Exclusion Argument to be invalid.

The modus tollens strategy is still available to my opponents for cases in which supervenience obtains. Thus, if a type identity theorist successfully argued that the Exclusion Principle were true, and the Exclusion Argument valid, he could then maintain that the irreducibility requirement must be false for those mental items, like qualia, that supervene on the physical. Therefore, to defend my position successfully, I must give independent reasons either for the falsity of the Exclusion Principle or for the invalidity of the Exclusion Argument.

In addition, the modus tollens strategy is also available for cases in which global supervenience obtains. If global supervenience, which is compatible with the most

plausible arguments against strong and weak supervenience, obtains, then one could still argue that another premise must be false, provided the Exclusion Argument is valid.

All of these considerations are essentially irrelevant to my strategy in the final part of this work. I do not pursue a strategy of applying modus tollens to the conditional corresponding to the Exclusion Argument. Instead I pursue the possibility of mental causation by giving independent reasons that this argument, even with a global supervenience thesis for content properties or a strong supervenience thesis for mental properties, is invalid. Thus, I will argue that mental causation is possible, if Nonreductive Physicalism is true, because the Exclusion Argument is invalid.

CHAPTER 8

THE EXCLUSION PRINCIPLE

8.1 Introduction

In previous chapters I have presented the Exclusion Argument against the possibility of mental causation and have argued that the conclusion of the Exclusion Argument could not be true. Thus, I concluded that the argument must be unsound. The premises that seemed most likely to be false were the Supervenience thesis and the Exclusion Principle. I argued in the previous chapter that the falsity of the Supervenience thesis did not provide a way for a dual-explanandum theory to take advantage of its falsity, and so failure of supervenience could not solve the problem of mental causation.

One might think that rejecting the supervenience thesis is sufficient for the Exclusion Argument to be unsound. Thus, there is no point in pursuing the argument any further. This reaction would be premature, however.

There are several reasons to continue pursuing the Exclusion Argument and for evaluating the Exclusion Principle. First, the best arguments in the previous chapter for rejecting the supervenience thesis only applied to content properties. The arguments for failure of supervenience based on absent qualia cases or quantum holism were not entirely convincing. Thus, mental properties, non-content properties, still appear to supervene on the physical. So, for those properties at least, we still need to find a flaw in the Exclusion Argument to show how it is possible for them to be causally relevant.

Second, global supervenience appears to obtain for mental content properties. Global supervenience may still suffice to make the Exclusion Argument valid even for content properties. The physical properties of a world fix the mental properties of that

world, so the mental may still be thought to play no causal role. Thus, we need to find some other flaw in the Exclusion Argument to allow for the possibility of mental causation.

Finally, even if the Exclusion Argument were admitted to be unsound, that would not remove the Exclusion Problem in general. Because the reasons for failure of supervenience could not provide a mechanism for mental causation (or causal relevance of content), the mental still seems to be excluded from causal relevance. There still seems to be nothing for the mental to do, barring implausible parallelist views, even if the mental fails to supervene on the physical. Thus, the mental still seems to be excluded from causal relevance. Thus, investigation of the Exclusion Principle and the reasoning from it is still necessary to show how mental causation is possible.

In this chapter I will argue that the Exclusion Principle is itself to blame for the apparent causal irrelevance of the mental. The Exclusion Principle is either false, if it is formulated in a naive way, or it makes the Exclusion Argument invalid, if formulated in a more sophisticated way. I will give reasons to reject the naive Exclusion Principle and for accepting a more sophisticated Exclusion Principle which is compatible with mental causation.

First, I will show how one simple version of the Exclusion Principle might be false by considering an argument by Stephen Yablo (1992). I will accept Yablo's counterexample but give reasons to reject his solution to the problem. Second, I will consider some additional evidence for the falsity of Yablo's Exclusion Principle. Third, I will argue for Kim's (1989) more sophisticated Exclusion Principle. Fourth, I will argue that several views of the relation of the mental to the physical are compatible with this

Exclusion Principle. Finally, I will consider some objections to my solution to the problem of mental causation.

8.2 Yablo on Mental Causation

Stephen Yablo (1992) offers an elegant solution to the problem of mental causation by arguing that the mental and the physical are related as determinable to determinate. If the mental and physical are related in this way, then they clearly do not compete for causal relevance. This solution to the problem of mental causation nicely captures and explains the most prevalent intuitions about the relation of the mental to the physical. However, his analysis is unlikely to be correct because, I shall argue, it does not naturally fit the facts about the mental and the physical. Yablo's theory could gerrymander a set of mental and physical properties or events that had the proper correlations to fit his theory of mental and physical as determinable to determinate, but these gerrymandered cases cannot provide evidence of a close fit between these two types of relations.

Two facts do not fit Yablo's theory of the relation of the mental to the physical. First, not all mental properties supervene on physical properties as Yablo's solution requires. Thus, his solution is incomplete and, if forced to take these cases into account, it is *ad hoc*. Second, the fact that some physical properties are outside the scope of mental properties contradicts Yablo's model of mental properties as determinables of physical determinates.

First, I will set up the problem of mental causation as it appears in the Exclusion Argument in the terms Yablo uses. Second, I will explain Yablo's solution to this problem in terms of the determinate/determinable relation. Third, I will argue that the determinable-determinate relation does not fit the relation of the mental to the physical.

8.2.1 The Problem of Mental Causation

The two modern variants of Nonreductive Physicalism both appear to lead to Epiphenomenalism. The first variant of Nonreductive Physicalism is the view that the mental is constituted by the physical but is neither type nor token-identical to it¹. Mental events, then, are distinct from physical events. Therefore, the problem for the Constitution Physicalist is one of whether mental events can be causes.

The second variant of Nonreductive Physicalism combines Token Physicalism with Type Dualism. This view is that the every event or state is physical, but that mental properties do not correspond one-to-one with physical properties. To take an oversimplified example, my pain right now is a C-fiber firing, but for something to be a pain is not for it to be a C-fiber firing because octopi or Martians might feel pain without having C-fibers. Mental types are not identical to physical types, but mental tokens are identical to physical tokens. Any particular event is physical, but no mental property of an event is identical to any physical property. So the second variety of Nonreductive Physicalism combines Token Physicalism with Property (or Type) Dualism. Both varieties of Nonreductive Physicalism involve a variety of dualism as well.

Each variety of dualism, or Nonreductive Physicalism, gives rise to a distinct variety of Epiphenomenalism. McLaughlin (1989) following Broad (1925), recall from chapter three, characterizes the two types of Epiphenomenalism as follows.

Type Epiphenomenalism (Type-E). (a) Events can be causes in virtue of falling under physical types, but (b) events cannot be causes in virtue of falling under mental types.

Token Epiphenomenalism (Token-E). (i) Physical events can cause mental events, but (ii) mental events cannot cause anything. (1989, 109-10)

¹ Because the mental and the physical have different modal properties, these philosophers think, the mental is constituted by the physical but is not identical to it. As an analogy, a statue, Goliath, is not identical to the physical substance that makes it up, call it Lump1, because Goliath could exist without one of his fingers whereas Lump1 could not. And Lump1 might never have been a statue, if it had been made into something else, but Goliath could not have been other than a statue. The Goliath-Lump1 example comes from Gibbard (1975) although Gibbard draws a different conclusion.

These two Epiphenomenalisms arise from the Exclusion Argument, a set of premises, for the most part constitutive of Nonreductive Physicalism, which lead to the conclusion that the mental cannot be causally relevant. I will present Yablo's version of the Exclusion Argument briefly because it calls attention to his objection to the Exclusion Principle. Yablo's (1992) version of the Exclusion Argument consists of the following premises².

- (1) If an event x is causally sufficient for an event y , then no event x^* distinct from x is causally relevant to y (*exclusion*).
- (2) For every physical event y , some physical event x is causally sufficient for y (*physical determinism*).³
- (3) For every physical event x and mental event x^* , x is distinct from x^* . (*dualism*).
- (4) So: for every physical event y , no mental event x^* is causally relevant to y (*epiphenomenalism*). (1992, 247-48)

This argument immediately entails that our minds are irrelevant to our actions. Moreover, it entails that our minds cannot be causally relevant at all. Yablo writes,

Every event z of whatever type is metaphysically necessitated by some underlying physical event y , whose causally sufficient physical antecedents are presumably sufficient for z as well. But then by the exclusion principle, z 's mental antecedents are irrelevant to its occurrence. So mental phenomena are *absolutely* causally inert. (1992, 248)

The argument proceeds as follows. Because of physical determinism and the Exclusion Principle, the mental cannot affect the physical. Since the physical fixes the mental, the mental is unnecessary for other mental events, and so, by the Exclusion Principle, the mental cannot be relevant to either the mental or the physical. The mental, then, is epiphenomenal or causally inert. Yablo's argument when given in terms of events,

² Yablo gives the argument in terms of events but notes that the argument could be given equally well in terms of properties *mutatis mutandis*.

³ As Yablo notes in a footnote, this principle is flawed if physical indeterminism is true, but whether the physical is deterministic does not seem the important point. The principle could be replaced with some other that does not assume determinism, for example, one that assumed that all physical probabilities were

entails that mental events cannot be causes, and so leads to Token-E (and Type-E⁴). The argument, when given in terms of properties, entails that no event occurs in virtue of any mental property of the cause, and so leads to Type-E. Yablo tries to avoid both varieties of Epiphenomenalism.

8.2.2 Yablo's Theory

Yablo proposes a model for the relationship of the mental and the physical that avoids the problem of mental causation. According to Yablo, the physical is the determinate of the mental's determinable. If the physical and mental are related as determinate to determinable, then, Yablo argues, the causal Exclusion Principle should not apply to them because determinates and determinables, intuitively, do not compete for causal influence.

Yablo characterizes the relation of determinate to determinable as follows,

- (Δ) P determines Q iff: for a thing to be P is for it to be Q, not simpliciter, but in a specific way. (1992, 252)

The easiest way to grasp this notion is by means of examples. *Crimson* is a determinate of the determinable *red*. *Red* is a determinate of the determinable *colored*. *Crimson* is a subcategory of the category *red*, but there is nothing added to the notion of *red* that results in *crimson*. To be crimson is not to be red and something else but to be red in a particular way. Determinables supervene asymmetrically on determinates⁵. Anything that is crimson is

fixed by antecedent physical events. Another candidate to replace determinism as the principle is the Causal Closure of the Physical, that any physical event that is caused has a physical cause.

⁴ Token-E entails Type-E but not vice versa. If no mental event can be a cause, then no event can be a cause in virtue of being a mental event. It is possible, on the other hand, if mental events are physical events, for the event to be a cause in virtue of being physical but not in virtue of being mental. McLaughlin (1989) discusses these points in greater detail.

⁵ It is interesting to note that determinables supervene strongly on determinates in Kim's (1987b) terms. I will assume that strong supervenience is the type of supervenience at issue. If one believes that the mental can *only* weakly supervene on the physical, then one would not be convinced by Yablo's subsumption of the mental-physical relation to the determinable-determinate relation. Furthermore, if the mental only globally supervenes on the physical, then Yablo's solution does not fit the actual mental-physical relation. I will assume, as does Yablo, that the supervenience in question is strong supervenience. However, I will

necessarily red but it is not the case that anything that is red is (necessarily) crimson.

Yablo notes that the relation of asymmetric determination is the common philosophical view of the relationship of the mental to the physical. The physical determines the mental, but, because of the multiple realizability of the mental on the physical, it is possible for some event to have a particular mental property without having a particular physical property. Pain can occur in different creatures that share no relevant brain properties. So, Yablo characterizes the asymmetric determination of the mental by the physical as,

(M) Necessarily, for every mental property M, and every physical property P which necessitates M, possibly something possesses M but not P. (1992, 255)

But (M) gives the same correlation of properties that the determinate-determinable relation does.

Determinates and determinables do not, intuitively, compete for causal relevance. Because the Exclusion Principle, as given by Yablo, rules out causal relevance of determinables, it must be false. If any event or property is sufficient for an effect, then no other event or property can be causally relevant. However, it is implausible that determinates rule out causal relevance for their determinables. For example, a stoplight is red and colored⁶. If I see a stop light and stop, the red of the stoplight is causally relevant to my stopping, but, according to the Exclusion Principle, the color of the light (the fact that it

argue that content does not supervene either strongly or weakly, so neither strong nor weak supervenience captures the mental-physical relation.

⁶Although identifying the red-token with the color-token may provide a reason the instances are both causally relevant, Yablo does not think the putative identity would solve the problem. He writes, "Imagine a glass which shatters if Ella sings at 70 decibels or more. Tonight, as it happens, she sang at 80 db, with predictable results. Although it was relevant to the glass's shattering that the volume was 80 db, it contributed nothing that it was *under* 90 db. Therefore, an efficacious determinate can have an irrelevant determinable." (Yablo, 1992, p. 259) Yablo thinks this example shows that the putative token identity does not make all determinables causally relevant because the determinable *under 90 db* is clearly not relevant to the glass's shattering even though its determinate *80 db* is causally relevant, so identifying

is colored) is not causally relevant. Yablo believes this result is obviously absurd; causal relevance of a determinate cannot exclude causal relevance of the determinable. Thus, he concludes, the Exclusion Principle is false since it excludes genuinely causally relevant properties. Moreover, any revision of the Exclusion Principle⁷ weak enough to allow both determinates and determinables to be causally relevant would not exclude mental properties. I will take up this last issue in a moment, but first I will argue that Yablo's theory of the relation of the mental to the physical is incorrect.

8.2.3 Criticism of Yablo

Yablo argues to the best explanation to support his theory of the relation of the mental to the physical. The evidence he gives fit into essentially two categories. First, he gives some reason to think the determinate-determinable relation fits our conception of the relation between the mental and the physical. The standard conception in philosophy of mind of the relation between the physical and the mental fits nicely with the metaphysical determination relation that obtains between determinate and determinable. Yablo believes this close fit between the two types of relations is not a coincidence. The second category of evidence is fecundity of the hypothesis. Specifically, as we have already seen, the hypothesis solves the problem of mental causation. These points, Yablo believes, are evidence in favor of his hypothesis.

We have already seen how Yablo's hypothesis solves the problem of mental

instances of determinates and determinables will not guarantee causal relevance for the determinable, according to Yablo.

⁷ Such as the more sophisticated statements of the Exclusion Principle given by Kim (1987a) and (1989a) Kim's version of the Exclusion Principle is that there can be no more than one complete and independent causally relevant property. On Yablo's view of the mental, the mental and physical would not be independent properties, so there is no reason to think Kim's principle is false. It is sufficiently weak, however, that one could not conclude from it that the mental is causally irrelevant. Therefore, Yablo's model for the mental-physical relation is sufficient to avoid the problem of mental causation as suggested by Kim's exclusion argument as well as the exclusion argument presented by Yablo.

causation. The first type of evidence, when considered in its entirety, however, counts against, rather than for, Yablo's hypothesis. The physical-mental relation does not fit the determinate-determinable relation for two reasons. First, not all mental properties supervene on the physical. Second, physical properties occur without corresponding mental properties which could not be the case if the mental were a determinable to the physical's determinate.

8.2.3.1 Content Properties Failure to Supervene

First, mental properties, and special science properties generally, do not always supervene on the physical properties of which, according to Yablo, they are determinables. One requirement of the determinable-determinate relation is that fixing the determinate fixes the determinable; determinates asymmetrically necessitate their determinables. So, anything that is scarlet must also be red. Not all mental properties are fixed by their physical properties in this way. Content properties, for example, depend on one's connection to the world and to the society of which one is a part. For example, my thought that "water is wet" is only a thought about water because of some causal/historical or social relations between my mental tokens of "water" and water in the external world or the society of which I am part. This relation cannot be captured by local physical properties of the brain. Because of this failure of content properties to supervene locally on physical properties, Yablo's hypothesis leaves out at least one type of mental properties. Thus, since only some mental properties supervene on the physical, Yablo would be unjustified in taking supervenience to be essential to the relation of the mental to the physical.

Yablo recognizes this problem and gives several responses. First, he notes that it is a mistake to think that the events or properties involved in the supervenience relation must be localized brain events that can occur in isolation from their neural context. This response is

adequate for properties, e. g. pain or itchiness, that are plausibly determined by the whole brain⁸. However, this response does not address the failure of content to supervene on local physical properties. Yablo's response to this failure has several parts. In a footnote, he writes,

So-called "wide content" mental events raise related but different problems which I don't discuss. Possibly they will have to be allowed as exceptions to the physical/mental determination thesis; in that case, the paper should be read as defending the causal potency of *other-than-* "wide content" mental events. Two remarks, though, to put this in perspective: First, it is controversial how often such events are genuinely efficacious, in particular because their "narrow" counterparts seem ordinarily to be more commensurate . . . with their supposed effects . . . Second, determination is only the most obvious of a number of intimate identity-like relations equally unresponsive of the " x_1 was sufficient, so x_2 was irrelevant" reflex." (1992, 271)

Yablo's response to the possibility of wide content events, e. g. my thought that water is wet, is four-fold. First, he implies a certain skepticism about the existence of these wide content properties. This essay is not the place to argue that content properties are wide. I will assume that that thesis has been well-established elsewhere.⁹

Second, he argues that his solution can still work for those properties that supervene even if not all mental properties do. This response is *ad hoc* given Yablo's need for a close fit between the mental-physical relation and the determinable-determinate relation. Essentially, Yablo's response here is to claim that his solution, which depends on the mental having a certain relation to the physical, works for all mental properties that do have that relation. Admittedly, content properties constitute a variety of mental properties, but I see no reason to think that content properties are distinct in any way that matters except insofar

⁸ Assuming even this much localization may be contentious. Proponents of the Representational Theory of the Mind often contend that phenomenal properties are individuated externally. Thus, they claim that phenomenal properties are not in the head any more than content properties are. See Lycan (1987), chapter

as they fail to supervene on the physical. To assume that their failure to supervene is in itself sufficient to justify a separate treatment of them is to beg the question.

Third, Yablo questions the causal relevance of “wide content” and asserts that “narrow content” is a better candidate for causal relevance. This work is, again, not the place for a detailed discussion of the arguments for and against the existence of so-called “narrow content,” but, whatever the merits of the notion of narrow content, Yablo cannot accept the standard reasons for believing in it. Therefore, Yablo cannot accept the existence of narrow content except as an *ad hoc* measure.

Fodor’s (1987) argument for the existence of narrow content can be put crudely in the following way.

- (5) At least some mental content is causally relevant.
- (6) All causal relevance is individualistic.
- (7) Therefore, at least some mental content is individualistic.

The conclusion is that some individualistic or “narrow” or local content must exist to be causally relevant. The nature of this narrow content remains unclear. In fact, one cannot even express any example of narrow content because any statement must be in a language and so must be wide. In any case, Yablo cannot accept Fodor’s assumption that the mental is causally relevant since Yablo is presenting an argument for the mental’s causal relevance. Yablo cannot appeal to narrow content since the notion of narrow content was *invented* to fit our concept of locality of causal relevance. Yablo cannot rely on a supervenient type of content to save his theory since that notion of narrow content was stipulated to be local. The only reason (or primary reason) to think narrow content exists is because it would provide the kind of local or supervenient property that fits our notion of only local causal relevance.

8 and (1996), chapters 4-5, and Tye (1995) for more discussion. I noted my reasons for rejecting this qualia externalism in the previous chapter.

But that assumption is closed to Yablo because he is arguing for a relation of the mental to the physical that is local, in which the mental supervenes on the physical. Appealing to narrow content assumes what Yablo must show, that the fit of mental to physical is that of determinable to determinate. Essentially, the notion of narrow content was constructed for the purpose of giving a supervenient type of content, so Yablo cannot appeal to it as an independently justified notion of content that conveniently fits his needs. Thus, Yablo cannot make use of the notion of narrow content except as an *ad hoc* measure to avoid the problem that not all mental properties supervene on the physical.

Yablo's fourth response, that the Exclusion Principle does not seem to apply to other identity-like relations, may be correct. However, unless one agrees with his other responses, this response provides a reason to *reject* Yablo's hypothesis about the relation of the mental to the physical in favor of one of these other identity-like relations. If the Exclusion Principle fails to apply to other identity-like relations, then one of them may be the actual relation of the mental to the physical.

8.2.3.2 Physical Properties without Corresponding Mental Properties

The second problem for Yablo's hypothesis is that some physical properties do not correspond to any mental property. This fact is inconsistent with the mental and physical being related as determinable to determinate. Determinables are inclusive of their determinates. According to Rosenberg,

[D]eterminables and determinates plainly differ in *scope*. Determinable properties are broader or more general than their corresponding determinates; determinate properties, narrower or more specific than their superordinate determinables. Determinables and determinates of a given family thus form a hierarchy of scope-inclusions, after the manner of *coloured* and *red*, *red* and *crimson*, and *crimson* and *Harvard crimson*. (1995, 116)

⁹ Seminal articles on wide content are Putnam (1973) and (1975) and Burge (1979).

So, there can be no object that is red or blue or yellow without that object being colored. No determinate can exist outside the scope of its determinable.

Contrary to Yablo's claim, then, the physical cannot be the determinate of the mental determinable. Yablo considers it possible that mental properties not correspond to any physical property. Thus, he thinks that physical properties are not inclusive of mental properties. That fact, if fact it is, supports Yablo's hypothesis. However, Yablo does not consider that some physical properties are outside the scope of the mental properties. Certain physical structures do not have any mental property associated with them. My C-fiber firing may correlate with pain, and some different neural firing may correlate with some other feeling, say an itch, but the physical structure of a rock clearly has no mental property associated with it. Examples of physical properties which fall outside the scope of the mental are embarrassingly easy to find. Rocks, chairs, tables, stars, and atoms all have physical properties but no correlative mental property. However, since determinables include their determinates, no determinate can fall outside the scope of its determinable. Since some physical properties clearly fall outside the scope of any mental properties, the physical cannot be a determinate of a mental determinable.

Perhaps this criticism is unfair to Yablo. For Yablo can argue that only some physical properties are determinates of their mental determinables, but not all physical properties are determinates of a mental determinable. Thus, my C-fiber firing would be a determinate of the determinable pain, but the physical structure of a rock is not a determinate of any mental determinable. Only some physical properties, like active brain states, are determinates of a mental determinable.

There are two problems with this response. First, the response is, once again, *ad*

hoc. It is possible that only some physical properties are determinates, but if Yablo wishes to argue from the close fit of the determinable-determinate relation to the mental-physical relation, he cannot include all and only those mental and physical properties that have that close fit. The evidence for Yablo's claim is that a close fit obtains between the mental-physical relation and the determinable-determinate relation. But if Yablo gerrymanders the pairs of properties that he considers in support of his thesis, then he has given up the reason he had for his theory. It is possible that some sets of properties meet his specifications, but only gerrymandered sets of properties do. Thus, Yablo no longer has good evidence for his thesis.

Second, the response overlooks the fact that the physical structure of the rock determines that the rock has no mental property. The physical features of the rock are as responsible for its lack of mental properties as my C-fiber firing is to my having a pain. The physical properties of an object still determine the mental properties, or lack of mental properties, of that object. Yablo's possible reply that only certain physical properties are determinates of mental determinables conspicuously fails to explain the fact that rocks and other purely physical objects lack mental properties. A theory of the mental ought to explain why brains have mental properties and other physical objects, like rocks, lack mental properties. A good theory of the mental ought to explain not just why some physical properties have mental correlates, but also why other physical properties do not. Yablo's theory, interpreted as claiming that some physical properties are determinates of mental determinables, does not explain why these other physical properties are not determinables of mental properties. There is one thing to be explained, why some physical properties correlate with mental properties and others do not. Yablo's theory can only explain why the

physical properties that do have a correlative mental property have one, but does not explain why these other physical properties lack correlative mental properties. Thus, although Yablo's theory may fit with some limited set of evidence of the relation of the mental and the physical, the totality of the evidence does not support his theory.

8.2.4 Concluding Remarks on Yablo

Yablo's model of the relation of the mental to the physical fails to accord with our ordinary conceptions and with some obvious facts. This evidence does not prove that the relation of the mental and physical is not one of determinable to determinate, but it undercuts his reason for drawing that conclusion. Yablo derives his conclusion only for a limited set of mental and physical properties. He provides an elegant solution to the problem of mental causation by hypothesizing that the mental is the determinable of the physical determinate. However, the evidence contradicts his hypothesis since the relation of the mental to the physical does not meet the requirements for Yablo's theorized relation. First, some mental properties do not supervene on physical properties, so Yablo's solution only works for those properties that do supervene. But this response is *ad hoc* and fails to give some relevant feature that might distinguish the supervenient from non supervenient mental properties in any non-question-begging way. Second, some physical properties are outside the scope of the mental which could not be the case if the mental were the determinable of the physical. Yablo could limit the set of physical properties to those which do correspond to a mental property, but that would also limit the evidence for his view in an *ad hoc* way. Yablo's view is not incoherent or inconsistent, but the set of evidence that supports his view is essentially gerrymandered to have the right fit. The evidence, overall, does not support his theory of the mental as a determinable of a physical determinate.

Yablo provides an excellent example, however, of a case in which the Exclusion Principle is false. He opens the door for other conceptions of the mental and the physical to provide counterexamples to the Exclusion Principle. Some close identity-like relation between the physical and the mental seems necessary to allow them not to compete for causal relevance. In the next section I will give a few more counterexamples to the naive Exclusion Principle. Then I will mention some theories of the mental that fit with rejection of Yablo's Exclusion Principle. Finally I will raise some problems that need to be addressed by my theory.

My solution to the problem presented by the Exclusion Argument is to reject Yablo's Exclusion Principle. If the Exclusion Principle is false, then the argument is unsound, and so it is possible for the mental to be causally relevant. If we accept a qualified or weakened Exclusion Principle, like the one advocated by Kim, that principle will not exclude the mental from causal relevance on one of a number of theories of the mental.

8.3 Rejecting the Exclusion Principle.

Aside from Yablo's counterexample, several other counterexamples to the Exclusion Principle have been given. These counterexamples are all cases in which correlated events, events related to each other either conceptually or nomically, are causally relevant. First, mediate causes of an effect are causes of that effect even though immediate causes of that effect are complete causes. Second, sometimes more than one property is relevant to an effect even if one alone is sufficient. The general lesson of these counterexamples is that correlated events or properties need not compete for causal relevance.

The first counterexample, from Goldman (1968) is that of mediate causes. A mediate cause, *C*, of some effect *A* is the immediate cause of some effect *B* which then

immediately causes *A*. In fact, a mediate cause could be related to an effect by an indefinite series of causal links. *B* is a complete cause of *A* in that no other cause is necessary for *B* to bring about the effect *A*. Imagine, for example, in a pool game that I strike the cue ball with a pool cue; the cue ball strikes the eight ball; and the eight ball goes into the corner pocket. My striking the cue ball causes the eight ball to go into the corner pocket; it is not ruled out from being a cause by the fact that the cue ball is the immediate cause of the eight ball going into the corner pocket. The Exclusion Principle, again, is the thesis that, if any effect has a complete cause, then it cannot have another cause. Since the immediate cause is a complete cause of the effect, the Exclusion Principle incorrectly excludes mediate causes as irrelevant when they are not.

Brian McLaughlin (1989) gives several examples involving properties that do not compete with each other for relevance¹⁰. Following the general spirit of his examples, we can construct a case which contradicts the Exclusion Principle. Imagine that Bill, who is married, attempts to marry again. His being married is causally relevant to his being arrested for bigamy, but his having a wife is also causally relevant to his being arrested for bigamy. These two properties, the property of being married and the property of having a wife, do not compete for causal relevance even though they are not identical properties.

Imagine a further example in which I try to hail a taxi by raising my arm. I raise my arm quickly to get the driver's attention. My raising my arm quickly gets the driver's attention, but does this mean that my raising my arm does not get the driver's attention? Presumably not. Neither the event of my raising my arm nor the event of my raising my arm quickly excludes the other from causal relevance. According to Yablo's Exclusion

¹⁰ This shift from events to properties addresses the possibility of Type-E rather than Token-E, but otherwise does not matter to the argument.

Principle, the two events must compete for causal relevance, but these conceptually related events do not exclude each other from causal relevance. Therefore, the naive Exclusion Principle is false for conceptually related events.

These cases, it should be emphasized, are not cases in which there is really only one event-type. No Exclusion Principle could apply to two identical events; showing that it did not apply would be trivial. But in this case, I am making the nontrivial claim that events that are conceptually or nomically related do not exclude each other. So, we need to note that the events described are not identical event-types¹¹. I could raise my arm without raising it quickly. The stoplight could turn red without turning exactly that shade of red. In either of these cases, counter to the Exclusion Principle, two closely related but non-identical events or properties are both causally relevant. Thus, we must reject the Exclusion Principle as it is stated above, that if there is a complete cause of an effect, then there can be no other cause of that effect.

8.4 A New Exclusion Principle

Suppose we now accept that the principle as stated is false. We can, however, state the Exclusion Principle in a way that makes it true (in the spirit of Kim, 1987a).

If there is one complete, *independent* cause of an effect, then there cannot be another cause for that effect.

This statement of the Exclusion Principle cannot be shown to be false by any of the counterexamples mentioned. Mediate and immediate causes of an effect are causally dependent; the other pairs of events or properties are conceptually dependent. So, the

¹¹ One might think these examples represent identical event-tokens but distinct event-types. It really makes no difference which view one holds. If the event-tokens are identical, then the distinction is whether the distinct properties of the single event compete for causal relevance. If the event tokens are distinct, if for example, I could perform exactly this arm-raising without doing it quickly, then the question is whether the distinct events compete as causes. All that is required to make the point is that the two events fail to be type-identical.

counterexamples to the Exclusion Principle as originally stated do not apply to the modified Exclusion Principle.

Having rejected one Exclusion Principle, one wonders whether one should accept any Exclusion Principle. Kim's (1987a) main argument for the Explanatory Exclusion Principle depends on his claim that having more than one explanation creates an epistemic tension.

- (8) Two purposes of explanatory practice are simplification and unification.
- (9) Giving multiple explanations of a single phenomenon cuts against the purpose of simplification.
- (10) Showing dependencies among different explanations better unifies the explanation.
- (11) Therefore, the practice of explanation should involve reducing multiple explanations to a single explanation and showing how the multiple explanations depend on each other.

This argument so far supports only the Explanatory Exclusion Principle, an epistemic principle not a metaphysical principle. Kim (1989a) argues further that this epistemic principle supports a general metaphysical principle. Assuming explanatory realism, we can see that an objective relation must hold between our theories and the world. If that is so, then the purposes of simplification and unification are taken as likely to lead to true explanations. If those explanations are true, then something about the world must correspond, in some way, with those explanations. So, the Explanatory Exclusion Principle, although itself only an epistemic principle, corresponds with a metaphysical Exclusion Principle. Kim takes this Exclusion Principle to be a Causal Exclusion Principle, but he recognizes that the principle must apply to any potential explanandum or any relation that grounds an explanation. So, as noted in chapter three, Kim can formulate a more general principle.

Generative/Determinative Exclusion Principle: there can be no more than one

complete, independent determiner of any event or property; if there is a complete, independent determiner of an event or property, then there is no other determiner of that event or property.

This more sophisticated Exclusion Principle is both broader and narrower than Yablo's Exclusion Principle. First, it is broader in that it applies both to effects and to properties of events. The mental properties of an effect are not, in any normal sense, caused by the physical properties of that effect or by the cause of the effect. The G/D Exclusion Principle is broader than Yablo's principle in that it includes properties as well as causes.

This more sophisticated Exclusion Principle is narrower as well, however, since it does not rule dependent phenomena to be causally irrelevant. This new Exclusion Principle does not exclude the mental from causal relevance, as long as the mental depends on the physical. Moreover, Kim's arguments support the less strict Exclusion Principle, the one that does not exclude mental causation. Thus, we can accept an Exclusion Principle that allows for mental causation and coheres with explanatory practice, while at the same time rejecting the stricter principle that is inconsistent with mental causation.

The Exclusion Argument with Yablo's Exclusion Principle is valid but unsound; the Exclusion Argument with the G/D Exclusion Principle is invalid. Yablo's Exclusion Principle is false since it incorrectly excludes conceptually or nomically related events as causally irrelevant. The G/D Exclusion Principle, on the other hand, does not exclude causes that are conceptually or nomically related to each other, and, if the mental is related to the physical in such a way, the Exclusion Argument does not exclude the mental from causal relevance. The Exclusion Argument in its schematic form, as presented in chapter 1, but with the replacement of the G/D Exclusion Principle for a more generic Exclusion Principle goes as follows.

- (12) *Dualism*. Mental and physical items are both real.
- (13) *Irreducibility*. No mental item is identical to any physical item.
- (14) *Supervenience*. Mental items supervene on or are determined by physical items.
- (15) *Causal Closure of the Physical*. No physical effect has a nonphysical cause.
- (16) *Generative/Determinative Exclusion Principle*: there can be no more than one complete, independent determiner of any item; if there is a complete, independent determiner of an item, then there is no other determiner of that item (except in cases of causal overdetermination).
- (17) Mental causation is not a case of causal overdetermination.
- (18) Therefore, *Epiphenomenalism*. No mental item causes any effect.

Mental to mental causation is compatible with these premises if the mental is not independent of the physical. The Causal Closure principle rules out the possibility of downwards causation, causation of the physical by the mental, but these premises allow for mental to mental causation (or quausation). In other words, a dual-explanandum theory is compatible with the Exclusion Argument. Thus, the Exclusion Argument is invalid. Even if these Physicalist assumptions are true, mental to mental causation is possible if the mental depends on the physical.

The schematic Exclusion Argument involves an oversimplification. As we saw in chapter three, more than one type of Causal Closure principle is possible, and not all such principles require mental causes to be causally irrelevant to physical events or properties. A brief reminder of the exclusive and nonexclusive types of Closure principles, from chapter three, may be helpful here.

The first two Closure principles are exclusive. They claim that there can be no mental cause of a physical effect. These principles would be rendered false if it were possible for both a mental and a physical cause to cause a physical event.

Causal Closure of the Physical:

Necessarily no non-physical event causes a physical event.

Causal Closure of Physical properties:

Necessarily no non-physical property is causally relevant to any physical effect or to any effect insofar as that effect is physical.

The exclusive Causal Closure principles may well be false given the counterexamples to Yablo's Exclusion Principle. Yablo's Exclusion Principle, we saw, was false, and, if we accept the counterexamples to Yablo's Exclusion Principle, then perhaps we should also reject the exclusive Closure principles. I can think of no direct evidence that would show the Exclusive Causal Closure principle to be false. The counterexamples to Yablo's Exclusion Principle all involved uncontroversial cases. The case of causal relevance of the mental to the physical is part of the question at issue, so we cannot assume that it is true or false without argument.

The safer assumption is to take a Nonexclusive Causal Closure Principle to be true. Arguments for the Closure principle, I noted in the introduction to this set of chapters, are fairly uncommon, but one argument often given by Kim (1993b and elsewhere) is that to deny Causal Closure of the Physical is to suppose some direct causation on the physical by a mysterious mental entity, an entelechy, by some kind of telekinesis. However, this argument does not give reason in support of the Exclusive Causal Closure Principles but only the Nonexclusive Causal Closure Principles. Nonexclusive Causal Closure Principles say that there can be no direct causation of the physical without there also being a physical cause. Nonexclusive Causal Closure Principles could be given as follows.

Nonexclusive Causal Closure of the Physical for events:

Necessarily, for any physical effect, there is a physical cause for that event.

Nonexclusive Causal Closure of the Physical for properties:

For any physical property of an effect, necessarily, there is a physical property of the cause that is relevant to the physical effect and to that effect insofar as it is physical.

These Causal Closure principles need not exclude the mental from relevance to the physical. Thus, rejecting this exclusive Closure principle, and accepting a weaker nonexclusive Causal Closure, which is all we are justified in supposing from Kim's argument, allows us to accept some theory other than a dual-explanandum view. If the mental can be relevant to the physical, then there is no need to postulate dual-explananda. We can suppose that a single explanandum sometimes has more than one cause or property causally relevant to its occurrence. As long as the events or properties are properly correlated, then they can both be causally relevant to a single effect.

Rejecting Yablo's Exclusion Principle and the Exclusive Causal Closure principles allows for more possible solutions to the problem of mental causation. Rejecting these principles allows for a single-explanandum solution to the problem of mental causation. However, explanatory and scientific practice still militate in favor of a dual-explanandum theory if one is to be had. Our practice generally explains events at one level by means of events at the same level. For example, behavior, beliefs and desires are generally explained in psychological or folk psychological terms. Sometimes we do give reductive explanations, or explanations of an event in terms of microstructure or lower-level implementation, but we rarely give explanations of physics in terms of the mental¹². So, a dual-explanandum view, although not necessitated by the Exclusion Argument, properly understood, still best coheres with our explanatory practice.

8.5 Mental Causation

Showing what is wrong with the exclusion argument is not itself sufficient to show how the mental can be causally relevant. How does such causation occur? Any theory that

¹² Excepting the integral use of measurement as part of quantum mechanics.

assumes the mental depends on the physical is compatible with the Generative/Determinative (hereafter G/D) Exclusion Principle. Thus, the G/D Exclusion Principle does not determine the correct theory of the mental-physical relation. Supporting a theory of the relation of the mental to the physical must appeal to more than just the Exclusion Argument, and canvassing reasons for different theories is outside the scope of this work. However, the Exclusion Argument does constrain the choice of possible theories to ones in which some close dependence relation holds between the mental and the physical.

Several common theories of the mind are compatible with the Exclusion Argument. One theory is that the physical realizes or implements the mental. The mental is, sometimes, conceived to be a second-order property of the physical realization. A second theory is that the mind is a macroscopic entity constituted by a physical brain. Thus, the mental mereologically supervenes on the physical. A third theory is that the mental 'emerges' from the physical given certain organizations of matter. On this view, the mental is a distinct entity from the physical but is dependent on the physical. Any of these views of the mental allows for mental causation compatible with the G/D Exclusion Principle. Any of a number of common Nonreductive Physicalist views that does not take context-dependence to be individuating of causally relevant properties is compatible with this G/D Exclusion Principle.

I would like to propose, somewhat tentatively, a Nonreductive Physicalist dual-explanandum theory of mental causation. On this view, the physical is the lower-level implementing mechanism for the mental. The mental causes behavior by means of a physical implementing mechanism. On this theory, my neural firing is not a behavior I do, but the neural firing is the physical implementation by means of which I behave. This

notion of the physical as an implementing mechanism is common in discussions of the mental, other special science properties, and about macroscopic events constituted by microevents. For example, we might say that the increased kinetic energy of the particles constituting a gas implemented the increase of pressure in that gas. Or we might say that the silicon chips and their operations implemented a computer's calculation. Thus, compatible with the G/D Exclusion Principle, a natural understanding of mental causation is that the mental causes mental or behavioral effects by means of a physical implementing mechanism.

For a dual-explanandum view of this sort to work the explananda must be distinct. I have already discussed reasons for irreducibility of the mental to the physical. Reasons for thinking behavior is irreducible to mere bodily movements are reminiscent of these reasons. One possibility is that behavior is not type identical to physical movements. Behavior is multiply realizable in different physical movements. For example, I could signal a left turn with different bodily movements. I could hit the turn signal on my car or I could put my arm out the window. Different physical movements might constitute the behavior of going to the fridge to get a beer. Lots of different physical movements could constitute beer-getting behavior because of differences in location of beer and differences in the person getting the beer. On this view, behavior may be token identical with the physical movements that constitute it, but the behavior type is not identical to any type of bodily movements. Thus, the mental could be relevant to the intentional act of behavior, to the physical movement's falling under that behavior-type. The mental causes other mental effects and behavior.

The cause insofar as it is mental, on this view, may be causally relevant to the effect

insofar as the effect is mental. Recall Horgan's (1989) neologism *quausion*. The mental is said to quause mental effects if and only if the effect is such a mental effect in virtue of the mental property of the cause. Accepting only the G/D Exclusion Principle, on a Token Identity view, allows the mental to quause other mental effects and behavior. Even though the mental property, M_1 , of the effect is determined, the mental property of the cause can still be relevant to M_1 .

On other views, behavior is not even token identical to the bodily movements that constitute it. Behavior has different properties than the physical movements constituting it. For example, my act of assassinating President Kennedy has different properties than the bodily movement or set of bodily movements that constitute it. This behavior has a property that no one physical movement I make has, namely the time during which the act occurs. My assassinating Kennedy does not occur at the time that I pull the trigger from my vantage point on the grassy knoll. Assassinating Kennedy requires that he be dead at the end of the behavior, and Kennedy does not die until after he has been taken to a hospital. So, my assassination of the President cannot be identical with my pulling the trigger. Thus, behavior cannot in general be identical to any single physical movement.

Moreover, behavior is not identical to any series of movements that constitute it. My assassinating the President could have differed in some details without ceasing to be that very behavior. For example, say that my assassinating the President is constituted by a series of events including my buying the gun, hiding on the grassy knoll, firing the shots, running away, and the events involved in the President dying. I could have done at least one of those bodily movements (ignoring for the sake of simplicity that some of these events are complex actions in their own right) differently and still have committed the very same action

of assassinating the President. But I could not have done any of these movements differently while these movements remained the same set of bodily movements. Thus, my behavior is not identical to any bodily movement or the set of bodily movements that make it up¹³.

Since behavior and the movements or sets of movements that constitute it are not identical, one can give a mental explanation of behavior without denying any premise of the Exclusion Argument. Since the mental is not identical to the physical, a mental cause of a mental effect does not also need to be a cause of a physical effect. Thus, the Causal Closure principles and the Exclusion Principle allow for mental to mental causation or mental quausation.

This theory is my solution to the problem of mental causation as it is raised by the Exclusion Argument. Let me note a constraint on my solution, and then I will deal with objections to the theory. The constraint is that a theory of mental causation cannot postulate new causal powers for the mental. Other typical dual-explanandum theories attempt to show how the mental adds some new, unique causal explanation of behavior. But this attempt violates the constraint on dual-explanandum theories that they cannot posit new causal powers for the mental without rejecting the dependence of the mental on the physical. The mental cannot have a new causal power unless it causes some events that are not already determined. Thus, finding new causal powers for the mental requires rejecting the dependence of the mental on the physical. And, as I argued in chapter seven, rejecting that premise does not provide an adequate theory of mental causation.

¹³ I have ignored the possibility that one might have a mixed set of views on irreducibility of behavior and the mental. For example, one might have a Token Physicalist view of behavior but a Constitution Physicalist view of mental events. These mixtures of views would complicate matters, but I believe the general solution I am proposing could be rephrased for those mixed views.

However, since my theory rejects Yablo's Exclusion Principle and accepts the G/D Exclusion Principle, I need not find any new causal power for the mental. Instead I can admit that the physical completely determines behavior, but insist that it is still possible that the mental determine behavior. Therefore, my theory of mental causation is compatible with the Exclusion Argument.

8.6 Objections

I would now like to raise some objections to my solution to the problem of mental causation and respond to them.

8.6.1 Novel Causal Powers

The first common objection is that rejecting the exclusion principle does not make room for the mental to have any new causal power. Since the physical already complete causes everything, one might ask, what does the mental add to the causal power the physical already has? In fact, I think this lack of new causal power is an advantage of my theory. As I have noted, the proper response is not to enumerate some new causal power of the mental. Instead the response is to diagnose the question as an erroneous way of thinking about the problem. The mental does not need to add any causal power to cause behavior. The counterexamples to the Exclusion Principle should be sufficient to convince us that the mental can be causally relevant as long as it depends on the physical. Requiring a new causal power for the mental is simply the wrong way to think about mental causation.

8.6.2 What God Needs to Do

Another objection along the same lines is that if there were no mental causation but only physical causation and determination of the mental by the physical, the world would be completely indistinguishable from the way it actually is. If mental causation does not make

any difference to the world, then it cannot be real. Or put another way, if God created a universe with only physical causation and determination relations, that world would not differ from our own. If our world does not differ from the world without mental causation, then we should believe ours is the simpler world, the world without mental causation.

Again, this way of thinking about the problem is misguided. Requiring the mental to make a noticeable difference in the world is a mistake given that the mental is dependent on the physical. I think this objection is reminiscent of the empiricist objection to the notion of causation. The empiricist objection is that regularities captured all there was to say about the world. What could causation add to the world? If God made all the regularities, he would not need to add causation. Causation, the argument goes, is unnecessary, and so we should never postulate it.

The proper response on the part of the realist about causation is to say that causation explains those regularities. The regularities supposed by the empiricist are not a brute unexplainable fact. Causation explains and grounds those regularities. The mental causation theorist can give the same answer. Mental causation in part, because it is part of a complete cause, explains the regularities. Mental causation is no more an unnecessary postulate than causation is.

One might object further that the physical explanation is all the explanation one needs. But mental causation is still part of that explanation. The physical explanation cannot be complete and independent without including the mental. Thus, this objection no more gives reason to reject mental causation than it does to reject causation. And I am confident that causation is a secure part of our ontology.

8.6.3 Epiphenomenal Causation

Another problem with the G/D Exclusion Principle is that it does not exclude epiphenomena. If epiphenomena are nomically related to or depend on an actual cause, then they are not excluded as causally irrelevant by this principle. For example, imagine that someone explains epileptic seizures as demonic possession. Neuroscience gives a better explanation in terms of uncontrolled neural firings. With Yablo's stricter Exclusion Principle, we could then reject the demonic possession theory. However, the clever demonic possession theorist could claim that the demons are nomically correlated with uncontrolled neural firings, and so his explanation is still correct, albeit incomplete in itself. So, the G/D Exclusion Principle is too weak to exclude entities that are genuinely causally irrelevant as long as they depend on causally relevant entities.

This argument differs from the previous objections in that it does not give a reason to think the Exclusion Principle is incorrect, but says that if the Exclusion Principle is as weak as I have argued, then it is impossible to exclude epiphenomena. The weakness of the Exclusion Principle does not show that we can give no reason to reject any entities as epiphenomenal. First, it still allows us to exclude epiphenomena if they do not depend on the physical. Thus, in the case of the demonic possession argument above, the demonic possession can be ruled out as a cause if it can be reasonably argued that demonic possession does not or cannot depend on brain activity.

However, the G/D Exclusion Principle does not exclude epiphenomena that really depend on the physical. For example, in the case of the whistle on a steam engine, we cannot apply the weaker Exclusion Principle to rule that the whistle does not cause the train's movement, provided the whistle nomically correlates with the train moving. How, then, do we determine whether some event really causes another?

First, if my arguments in the second section of this work, chapters four through six, are correct, then there really are no epiphenomenal particulars. My arguments in that section were that the mental cannot be epiphenomenal, but it seems that most of those arguments could apply to any particular property or existent. If those arguments are correct, then the real question is not whether an existent is causally relevant *simpliciter* but causally relevant to a particular effect. We still need to determine whether a cause is relevant to some effect, but we need not worry about whether an existent can be a cause at all. Noting this fact does not solve the problem but it does to some extent ameliorate it.

The response to this objection has to be, I think, that we should not expect the Exclusion Principle to exclude everything that is causally irrelevant. As an epistemic principle, the Exclusion Principle is imperfect, it does not give a sufficient condition for excluding epiphenomena that depend on genuine causes. It may be that sometimes we will not be able to distinguish an epiphenomenon from a genuine cause, but limitations on our ability to know should not be surprising. The G/D Exclusion Principle cannot rule out dependent epiphenomena as causes even assuming we have all the possible information.

Perhaps no single principle determines whether some entity is causally relevant, but some relations may ground genuine causes and others may not, and as a matter of common practice people are able to determine which cases ground genuine causation and which relations only involve epiphenomena. An appeal to common practice here is not inappropriate since we have already answered the questions of whether the mental is causally relevant (by arguing that Epiphenomenalism is incorrect) and how it might be causally relevant (by rejecting the overly strict Exclusion Principle). All the appeal to common practice does here is guess how one might in practice discern which entities are

genuine causes and which are not. We could wish for a principle that excluded epiphenomena in all cases, but perhaps no such principle would be possible.

8.6.4 Second Class Causation

A third objection is that models of mental causation like mine make mental causation to a second class status. It is alleged that mental causation, on this theory, is less real than physical causation or perhaps not real causation at all. Kim's (1984) supervenient causation is such a view that appears to relegate mental causation to a second-class status. According to Kim, for something to be a supervenient cause is for it to be "a causal relation that is reducible to, or explainable by, the causal processes taking place at a more basic physical level." (Kim 1984, reprinted 1993a, 107) If mental causation is reducible to or explainable by some more basic causation, then Kim's theory seems to make mental causation a second class citizen. Further, this notion of second-class causation does not seem to make sense. The mental can either cause or it cannot, and there is no room for anything else.

Although Kim has recently abandoned his theory of supervenient causation, the theory is worth discussing since it will show how my theory can avoid the same accusations. Kim clearly wants the supervenient cause to be a genuine cause. However, for an apparent cause to be explainable in terms of more basic physical processes is for that relation to be epiphenomenal, for the more basic relations to be the real cause. Put another way, nothing in Kim's theory will distinguish a supervenient cause from an epiphenomenon. For something to be a supervenient cause is for it to have only the causal powers of the more basic level. And that is what it is for something to be an epiphenomenon, to have no causal power of its own but only the causal power of the events that correlate with it. Supervenient

causation appears to be chimerical, a relation that correlates with the genuine physical causation but is not itself not real. So, Kim's notion of supervenient causation does not appear to give real causation, and, to the extent that my view takes advantage of the same Exclusion Principle for mental causation, my view appears to have the same problem.

One certainly does not want to claim that supervenient causes are some second-class variety of cause. If mental causation is to occur, then it must be as real as any causation. If the mental does not cause, then it is just a byproduct, an epiphenomenon. So, for my solution to allow for mental causation, mental causation must be real.

However, I think the question is misguided in this case. Recall some of the counterexamples to Yablo's Exclusion Principle. The redness of the light is, we think, causally relevant to one's stopping, but we have no inclination to ask whether the color of the light was also causally relevant. The color really is causally relevant, and it is not a second-class cause even though it depends on the redness of the light to exist. Since the mental depends on the physical in some sufficiently close way, the mental can be causally relevant just as the color of the stoplight is. The temptation to distinguish these properties and demand that only one be causally relevant is misguided. The mental can be causally relevant just as events dependent on more basic events can be.

Finally, one might ask whether my solution requires something like Kim's (1992 and 1993) Causal Inheritance Principle. Kim argues that because the mental depends on the physical, the mental must inherit its causal power from the physical. Otherwise, the mental's causal power emerges magically from nowhere. For example, on Kim's view, if my instance *pain* is to have an effect on my removing my hand from the stove, that instance of pain must gain its causal power from the physical implementing mechanism. The causal

power of this pain-token must be identical to the causal powers, or some subset of the causal powers, of its physical instantiation. Again, if the mental inherits its causal power from the physical, then the mental seems relegated to second class causal status.

My first inclination is to resist the *CIP*, but on careful consideration, I think that it is not unreasonable as a claim about instances. Certainly on a token-identity view, each mental token is no more than a physical token, and so the mental token can have only the causal powers that the physical token it is identical to has. As a theory about types, *CIP* would be deficient, but as a theory about tokens, as Kim intends it to be, it can still be true.

On Constitution Physicalism, the mental token is not identical to the physical token, and so it is questionable whether the mental token's causal powers must be identical to the physical token's. However, since we accept mereological supervenience, we must accept that the mental token has no new causal power lacked by the physical. So, on either token Physicalism or Constitution Physicalism, instances of mental events can have only the causal powers of their underlying physical instances.

Kim's metaphor of inheritance may be misleading, however. Nothing in the notion of dependence of mental causal powers on the physical requires that mental causation have a second-class status or that it be less real than physical causation. This way of thinking is the same type of error that would lead one to think that because the mental depends on the physical, the mental is somehow a second class entity or is less real than the physical. My pain's causal powers are no less real than my pain itself is. The conclusion that the mental has a second class status is a *nonsequitur*. It simply is not the case that the dependence of one entity on another makes the first entity less real than the second, whether these entities be existents, properties or causal powers.

The main reason I can see for resisting the notion of the *CIP* is that it seems as though some mental entities and their causal powers are not exhausted by the physical. The mental and other higher level properties often appear to be context-dependent. And this fact appears to give these entities unique causal powers that are not fully captured by the physical. For example, economic entities like money have certain causal powers given their role in society. Money seems to have unique causal powers, as described by the laws of economics, that go beyond the causal powers of the disjunctive physical realizations of money, say green pieces of paper, credit cards, etc. If context-dependence is an ubiquitous feature of the world, then it is important to explain the causal powers of context-dependent entities. I will address this objection separately.

8.6.5 Extrinsic Causal Relevance

The final objection is that my solution's requirement of dependence of the mental on the physical appears incompatible with causal relevance of contextually individuated entities. Mental content, as I discussed in the previous chapter, does not supervene on the physical, and so, there is no room for it to be causally relevant on this theory of mental causation. Money, on this theory, since it is context-dependent, does not have the close relation to the physical for it to be causally relevant. Yet, it seems obvious that money is causally relevant. Thus, my theory of mental and higher order causation must be false. Similar examples of mental content appear to show that my theory cannot work. The causal relevance of context-dependent entities seems to show that a dependence relation cannot be necessary for causal relevance.

My response to this objection has several parts. First, we should note that the problem of mental causation of the Exclusion Argument is distinct from the problem of

mental causation for extrinsic properties. The problem as presented in this work has been focused on the Exclusion Argument, an argument which presents the problem of mental causation only for those properties or entities that supervene on the physical. So, it is not necessary that this theory of mental causation also solve the distinct problem of mental causation for extrinsic properties.

However, one does not want to present a theory of mental causation that solves the problem only for a single type of problem and precludes a solution to other problems. Moreover, one does not want to provide a disjunctive solution for similar types. That is, if the mental is a single type of entity, providing a solution to the problem of mental causation for only some mental entities but not others is still inadequate.

Accepting the weak G/D Exclusion Principle may still allow for mental content to be causally relevant. Mental content does seem to depend on some complex physical relations. Worlds that are indiscernible with respect to all their physical properties and relations are indiscernible with respect to content properties. These physical relations are extrinsic or relational properties, but, presumably, causally relevant. Explaining how these extrinsic properties are causally relevant is a matter of solving the distinct problem of mental causation for extrinsic properties. So, assuming only the G/D Exclusion Principle, mental content that depends on these complex physical relations can be causally relevant as well. The Exclusion Principle only excludes entities that do not depend on the physical, so mental content may still be causally relevant on this account. Given rejection of Yablo's overly restrictive Exclusion Principle, one need not espouse an identity of content-types with physical relation-types; one need only claim that the content-type depends in the right way on these physical relation-types.

One final note on this problem with my solution to the problem of mental causation is that my solution works better on this count than a type identity theory does. One of the main thrusts of this work has been to solve the problem of mental causation for Nonreductive Physicalism. A return of type-identity theories has been fueled in large part by the problem of mental causation¹⁴. The objection that my view of mental causation is incompatible with causal relevance of extrinsic or content properties applies equally well to the type-identity theorist. So the type identity theorist cannot use this objection as fodder against the Nonreductive Physicalist. Thus, the goal of this work to defend Nonreductive Physicalism against the "New Reductionism" has been largely successful. My theory of mental causation for events or properties that supervene on the physical deals at least as well with extrinsic mental content as a type identity theory does.

8.7 Conclusion

Mental causation is consistent with Nonreductive Physicalism since the Exclusion Argument is invalid. The Exclusion Principle itself, when properly understood, does not exclude the mental from causal relevance given that the mental depends on the physical. The Exclusion Argument¹⁵ presents a problem of mental causation that requires us to reject some apparently plausible broadly physicalist assumption or to find a close relation of the mental and physical to which the Exclusion Principle does not apply. Rejecting Yablo's Exclusion Principle is a well-motivated way of solving this problem of mental causation. Moreover, rejecting this principle allows us to say how mental causation might occur. The counterexamples to Yablo's principle lead to a more sophisticated Exclusion Principle that does not exclude the mental from causal relevance. Adopting this sophisticated Exclusion

¹⁴ See especially Kim (1993b)

¹⁵ Perhaps the problem should just be called the Exclusion Problem at this point since the Exclusion

Principle does not necessarily provide everything we could wish for in a solution. It does not allow us to exclude epiphenomena. But it solves the problem of the Exclusion Argument without abandoning any basic precepts of Physicalism and without positing mysterious downwards causation of the mind without any physical mechanism. Finally, it avoids a return to type-identity theories that are inconsistent with multiple realizability and the Nonreductive Physicalist consensus in recent philosophy of mind. As I have tried to show, the nonreductivist has resources to deal with the problem of mental causation that are equal to or better than those available to the type identity theorist.

Argument may be unsound for reasons explored in the previous chapter.

WORKS CITED

- Adams, Robert Merrihew. 1986. Time and Thisness. *Midwest Studies in Philosophy*. 11: 315-29.
- Alexander, Samuel. 1927. *Space, Time, and Deity*, ii, 2d ed. London, Macmillan.
- Anscombe, G. E. M. 1963. *Intention*, 2d ed. Oxford: Basil Blackwell.
- Antony, Louise. 1989. Anomalous Monism and the Problem of Explanatory Force. *Philosophical Review* 2: 153-87.
- Baker, Lynne. 1993. Metaphysics and Mental Causation. In Heil and Mele 1993. 75-95.
- Bickle, John. 1992. Multiple Realizability and Psychophysical Reduction. *Behavior and Philosophy* 20: 47-58.
- Bieri, Peter. 1992. Trying Out Epiphenomenalism. *Erkenntnis* 36: 283-309.
- Block, Ned. 1978. Troubles with Functionalism. In *Perception and Cognition. Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science*, ed. C. W. Savage. 261-325. Minneapolis: Minnesota University Press.
- _____. 1990. Can the Mind Change the World? In *Meaning and Method: Essays in Honor of Hilary Putnam*, ed. George Boolos. 137-70. Cambridge: Cambridge University Press.
- Broad, C. D. 1925. *The Mind and Its Place in Nature*. London: K. Paul, Trench, Trubner & co. ltd.
- Burge, Tyler. 1979. Individualism and the Mental. In *Midwest Studies in Philosophy*. ed. P. French et al. 4: 73-121. Minneapolis: University of Minnesota Press.
- _____. 1993. Mind-Body Causation and Explanatory Practice. In Heil and Mele 1993. 97-120.
- Campbell, Keith. 1970. *Body and Mind*. Garden City, New York: Anchor books.
- Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Child, William. 1994. *Causality, Interpretation and Mind*. Oxford: Clarendon press.
- Churchland, Paul. 1984. *Matter and Consciousness*. Cambridge, MA: MIT Press.
- _____. 1989. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge: MIT Press.
- Davidson, Donald. 1963. Actions, Reasons and Causes. *Journal of Philosophy*. 60: 685-700.
- _____. 1970. Mental Events. In *Experience and Theory*. eds. L. Foster and J. Swanson. 79-101. Amherst, MA: University of Massachusetts Press.
- _____. 1969. The Individuation of Events. In *Essays on Actions and Events*. 216-234. Oxford: Oxford University Press, 1980. Originally published in *Essays in Honor of Carl G. Hempel*. ed. N. Rescher et al. Dordrecht: D. Reidel, 1969.
- _____. 1980. *Essays on Actions and Events*. New York: Oxford University Press.
- _____. 1986. Knowing One's Own Mind. *Proceedings of the American Philosophical Association* 60: 441-58.
- Dennett, Daniel. 1994. Real Consciousness. In *Consciousness in Philosophy and Cognitive Neuroscience*. eds. A. Revonsuo and M. Kamppinen. 55-63. Hillsdale, NJ: Lawrence Erlbaum.

- Descartes, Rene. 1984. *The Philosophical Writings of Descartes*. 3 vol. trans. John Cottingham, Robert Stuthoff, and Dugald Murdoch. Cambridge: Cambridge University Press.
- Dretske, Fred. 1988. *Explaining Behavior*. Cambridge, Mass.: Bradford Books, MIT Press.
- _____. 1989. Reasons and Causes. *Philosophical Perspectives*. 3: 1-15.
- _____. 1995. Reply: Causal Relevance and Explanatory Exclusion. In *Philosophy of Psychology: Debates on Psychological Explanation*. ed. Graeme and Cynthia MacDonald. 143-51. Oxford: Blackwell.
- Eccles, John. 1994. *How the Self Controls its Brain*. Berlin: Springer Verlag.
- Fodor, Jerry. 1974. Special Sciences, or the Disunity of Science as a Working Hypothesis. *Synthese* 28: 97-115.
- _____. 1987 *Psychosemantics*. Cambridge, Mass: MIT Press.
- _____. 1989. Making Mind Matter More. *Philosophical Topics*. 17: 59-79.
- Gassendi, Pierre. 1984. Fifth Set of Objections. In Descartes 1984. vol. 2. 179-240.
- Gibbard, Alan. 1975. Contingent Identity. *Journal of Philosophical Logic*. 4: 187-221.
- Goldman, Alvin. 1967. A Causal Theory of Knowing. *The Journal of Philosophy* 64: 357-72.
- _____. 1969. The Compatibility of Mechanism and Purpose. *The Philosophical Review* 78: 468-82.
- _____. 1970. *A Theory of Human Action*. Englewood Cliffs: Prentice-Hall.
- Gopnik, Alison. 1996. How We Know Our Own Minds: The Illusion of First-Person Knowledge of Intentionality. *Behavioral and Brain Sciences*. 16: 1-14.
- Grimes, Thomas. 1988. The Myth of Supervenience. *Pacific Philosophical Quarterly*. 69: 152-60.
- Hannan, Barbara. 1994. *Subjectivity and Reduction: An Introduction to the Mind-Body Problem*. Boulder: Westview Press.
- Hardin, C. L. 1990. Perception and Illusion. In *Mind and Cognition: A Reader*. ed. William Lycan. 555-67. Oxford: Basil Blackwell.
- Heil, John and Mele, Alfred. 1991. Mental Causes. *American Philosophical Quarterly*. 28: 61-71.
- _____, eds. 1993. *Mental Causation*. Oxford: Clarendon Press.
- Hellman, Geoffrey and Thompson, Frank. 1975. Ontology, Physicalism and Reduction. *Journal of Philosophy*. 72: 551-64.
- Honderich, Ted. 1982. The Argument for Anomalous Monism. *Analysis*. 42: 59-64.
- Horgan, Terence. 1989. Mental Causation. *Philosophical Perspectives*. 3: 47-76.
- Horowitz, Amir. 1999. Is There a Problem in Physicalist Epiphenomenalism? *Philosophy and Phenomenological Research*. 59: 421-34.
- Huxley, Thomas Henry. 1898. On the Hypothesis that Animals are Automata and its History. In *Collected Essays*. 4: 199-250. New York: Appleton and Company.
- Jackson, Frank. 1982. Epiphenomenal Qualia. *Philosophical Quarterly*. 32: 127-36.
- Johnston, Mark. 1985. Why Having a Mind Matters. In *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. eds. Ernest LePore and Brian McLaughlin. 408-26. Oxford: Basil Blackwell.

- Kim, Jaegwon. 1976. Events as Property Exemplifications. In *Action Theory*. ed. Myles Brand and Douglas Walton. 159-77. Dordrecht: D. Reidel Publishing Company. Reprinted in Kim 1993a.
- _____. 1984a. Concepts of Supervenience. *Philosophy and Phenomenological Research*. 45: 153-76. Reprinted in Kim 1993a.
- _____. 1984b. Epiphenomenal and Supervenient Causation. *Midwest Studies in Philosophy*. 9: 257-70. Reprinted in Kim 1993a.
- _____. 1987a. Explanatory Realism, Causal Realism, and Explanatory Exclusion. *Midwest Studies in Philosophy*. 12: 225-39.
- _____. 1987b. "Strong" and "Global" Supervenience Revisited. *Philosophy and Phenomenological Research*. 48: 315-26. Reprinted in Kim 1993a.
- _____. 1989a. Mechanism, Purpose and Explanatory Exclusion. *Philosophical Perspectives*. 3: 77-108. Reprinted in Kim 1993a.
- _____. 1989b. The Myth of Nonreductive Materialism. *Proceedings of the American Philosophical Association*. 63: 31-47. Reprinted in Kim 1993a.
- _____. 1990. Supervenience as a Philosophical Concept. *Metaphilosophy*. 21: 1-27.
- _____. 1991. Dretske on How Reasons Explain Behavior. In *Dretske and His Critics*. ed. Brian McLaughlin. 52-72. Oxford: Basil Blackwell. Reprinted in Kim 1993a.
- _____. 1992. Multiple Realizability and the Metaphysics of Reduction. *Philosophy and Phenomenological Research*. 52: 1-26. Reprinted in Kim 1993a.
- _____. 1993a. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.
- _____. 1993b. The Nonreductivist's Troubles with Mental Causation. In Heil and Mele 1993. 189-210. Reprinted in Kim 1993a.
- _____. 1995. Mental Causation: What? Me Worry? *Philosophical Issues*. 6: 123-51.
- _____. 1996. *Philosophy of Mind*, Boulder, CO: Westview Press.
- _____. 1997. Does the Problem of Mental Causation Generalize? *Proceedings of the Aristotelian Society*. New Series. 97: 281-97.
- _____. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, Mass: MIT Press, Bradford.
- Kripke, Saul. 1972. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lashley, Karl. 1956. Cerebral Organization and Behavior. In *The Brain and Human Behavior*. ed. H. Solomon, S. Cobb and W. Penfield. 1-18. Baltimore: Williams & Wilkins.
- LePore, Ernest and Loewer, Barry. 1987. Mind Matters. *Journal of Philosophy*. 93: 630-42.
- Levine, Joseph. 1983. Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly*. 64: 354-61.
- Lycan, William. 1987. *Consciousness*. Cambridge, Mass.: MIT Press.
- _____. 1996. *Consciousness and Experience*. Cambridge, Mass.: MIT Press.
- Malcolm, Norman. 1968. The Conceivability of Mechanism. *The Philosophical Review*. 77: 45-72.

- McLaughlin, Brian. 1989. Type Epiphenomenalism, Type Dualism and the Causal Priority of the Physical. *Philosophical Perspectives*. 3: 109-35.
- . 1993. On Davidson's Response to the Charge of Epiphenomenalism. In Heil and Mele 1993. 27-40.
- . 1994. Epiphenomenalism. *A Companion to the Philosophy of Mind*. ed. Samuel Guttenplan. 277-88. Oxford: Blackwell.
- Menzies, Peter. 1988. Against Causal Reductionism. *Mind*. 97: 551-74.
- Nagel, Thomas. 1974. What Is It Like to Be a Bat? *Philosophical Review*. 83: 435-50.
- Nisbett, Richard and Wilson, Timothy. 1977. Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*. 84: 231-59.
- Penrose, Roger. 1989. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. New York: Penguin Books.
- Pereboom, Derk and Kornblith, Hilary. 1991. The Metaphysics of Irreducibility. *Philosophical Studies*. 63: 125-45.
- Plato. 1989. *Phaedo* in *The Collected Dialogues of Plato, Including the Letters*, ed. Edith Hamilton and Huntington Cairns. Princeton, NJ: Princeton University press.
- Popper, Karl and Eccles, John. 1977. *The Self and its Brain*. New York: Springer International.
- Putnam, Hilary. 1967. Psychological Predicates. In *Art, Mind and Religion*, ed. W. H. Capitan and D. D. Merrill. 37-48. Pittsburgh, University of Pittsburgh Press.
- . 1973. Meaning and Reference. *Journal of Philosophy*. 70: 699-711.
- . 1975. The Meaning of "Meaning". *Language, Mind, and Knowledge*, ed. Keith Gunderson. 131-93. Minneapolis: University of Minnesota Press.
- Rosenberg, Jay. 1995. Determinate/determinable. *A Companion to Metaphysics*. ed. Jaegwon Kim and Ernest Sosa. 115-17. Oxford: Blackwell.
- Searle, John. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Sosa, Ernest. 1984. Mind-Body Interaction and Supervenient Causation. *Midwest Studies in Philosophy*. 9: 271-81.
- Shoemaker, Sydney. 1975. Functionalism and Qualia. *Philosophical Studies*. 27: 291-315.
- Smullyan, Raymond. 1980. An Unfortunate Dualist. In *This Book Needs No Title*. Englewood Cliffs: Prentice Hall. Reprinted in *The Mind's I*. 383-84. Toronto: Bantam Books, 1982.
- Stoutland, Frederick. 1985. Davidson on Intentional Behavior. In *Essays on Wittgenstein in Honor of G. H. Von Wright (Acta Philosophica Fennica, 28)*. 286-325.
- Tye, Michael. 1995. *Ten Problems of Consciousness*. Cambridge, Mass.: MIT Press.
- Van Gulick, Robert. 1992. Three Bad Arguments for Intentional Property Epiphenomenalism. *Erkenntnis*. 36: 311-31.
- . 1993. Who's in Charge Here? And Who's Doing All the Work. In Heil and Mele 1993. 233-256.
- Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. 3d ed. trans. G. E. M. Anscombe. New York: MacMillan.
- . 1960. *The Blue Book*. New York: Harper and Row.
- Yablo, Stephen. 1992. Mental Causation. *The Philosophical Review*. 101: 245-80.