PHYSICALISM AND EPIPHENOMENAL PROPERTIES

PHYSICALISM AND THE CHALLENGE OF EPIPHENOMENAL PROPERTIES

By

NEIL CAMPBELL, M.A.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (1997) McMaster University
(Philosophy) Hamilton, Ontario

TITLE: Physicalism and the Challenge of Epiphenomenal Properties.

AUTHOR: Neil Campbell, M.A. (McMaster University)

SUPERVISOR: Professor Evan Simpson

NUMBER OF PAGES: v, 149

ABSTRACT

The following dissertation is an examination of arguments against physicalism. Physicalism is a thesis in the philosophy of mind that is constituted by two central claims: (1) the ontological claim that everything that exists is ontologically physical and that human beings are among such things; (2) the explanatory claim that all facts about human beings and all explanations of their behaviour are dependent on and determined by physical facts and explanations. It has frequently been asserted that there are properties that escape capture in physicalist accounts of human behaviour, thereby undermining (2). Such properties are usually thought to be lacking causal powers, and hence have been called "epiphenomenal." The epiphenomenalist objections have long been thought to represent a serious obstacle to physicalism. My aim is to show that the objections that are motivated by epiphenomenal properties are unconvincing.

My discussion proceeds in two stages. In the first stage I examine the epiphenomenalist objections in detail and show that in their most persuasive forms they demonstrate that physicalism has certain explanatory inadequacies. The critics of physicalism believe that these shortcomings lead to the denial of the explanatory completeness of physicalism, and I try to make their case as charitably as I can. In the second stage of the argument I invoke the relation of psycho-physical supervenience and show that the desired conclusion does not follow, even if we admit that physicalism has

certain explanatory failings. The overall conclusion of this dissertation is that the epiphenomenalist objections to physicalism are completely undermined and hence that properties which were thought to be epiphenomenal do not represent a serious obstacle to physicalism as was previously thought. My intention is that this discussion push forward work in the philosophy of mind and point the way to a more adequate articulation of physicalism.

Acknowledgements

I would like to express my gratitude to those who have served on my thesis committee for their insightful comments, guidance, and support. This is especially true in the case of Evan Simpson, who has taken a great deal of care and thought in his criticisms of earlier drafts of this thesis, and for which I am extremely grateful. My work is much more precise and has taken more interesting directions as a result of Evan's suggestions. I would also like to thank Evan for the opportunity to serve as his research assistant in 1996. My work with Evan on the course in the philosophy of mind provided the occasion to develop some common philosophical ground between us and was very helpful to my research.

I would also like to express my thanks to a very special circle of friends who have made my research at McMaster one of the most enjoyable experiences of my life. You all know who you are. Thanks also to my parents Janis Hoogstraten and Colin Campbell, and my adopted "Aunties" Jane and Sandy for their encouragement and all of those Thai dinners when the writing got to be too much. Special thanks to Stan Clarke, a dear friend and mentor. And thanks also to Melissa Otto, who happily entered my life toward the end of my research and played a significant role in sustaining my sanity throughout the stresses associated with the final stages of my Ph.D. Finally, I wish to thank the Social Sciences and Humanities Research Council for their financial support which made a tremendous difference to my ability to finish my research on time.

Table of Contents

Chapter 1

Introduction

The "mind-body problem" as such is no longer of great concern to contemporary philosophers. I suspect this is due to the growing consensus that some form of physicalism is true, and thus that questions about ontology are no longer of great interest.[1] However, this does not mean that philosophers have stopped worrying about the mind. On the contrary, the number of recent books and articles by philosophers on consciousness indicates that the "philosophy of mind" remains at the forefront of Anglo-American analytic philosophy. Since most philosophers are convinced of the truth of physicalism much work in the philosophy of mind is concerned with working out the details of how conscious mental phenomena are physically realized. Generally speaking, among physicalists there are two attitudes toward this endeavor. On one side there are theories that claim conscious mental states can be incorporated into a broadly physical explanatory framework, while on the other side there are theories that deny this is possible. The first group is constituted by theories that show that mental states are identical with physical states of the brain or body (Type-Identity,[2] Token-Identity,[3] and with some qualifications, Functionalism[4]). The

---

[1] For an example of this attitude see William Seager, Metaphysics of Consciousness, Chapter 1. (New York: Routledge, 1991).

[2] U. T. Place, "Is Consciousness a Brain Process?" British Journal of Psychology 47 (1956) pp. 44-50; J. J. C. Smart, "Sensations and Brain Processes," Philosophical Review 68 (1959) pp. 141-156.

[3] Donald Davidson, (1970b) "Mental Events," in Davidson Essays on Actions and Events (Oxford: Oxford University Press, 1980).

[4] David Lewis, (1978) "Mad Pain and Martian Pain," in Readings in Philosophy of Psychology Vol. 1, ed. Ned Block (Cambridge: Harvard University Press, 1980).

second group, which denies there can be a physical understanding of the mind, divides into two categories. First, there are the proponents of Eliminative Materialism.[5] They claim there is nothing to understand about the relationship between mental phenomena and the physical sciences because there are no such things as mental phenomena. Beliefs, desires and other mental states are theoretical entities belonging to a theory of behaviour known as "folk psychology." Since, in their view, folk psychology is a false and misleading theory of human behaviour its theoretical posits will be eliminated along with the theory when it is eventually replaced by a more accurate scientific account of behaviour. Therefore, rather than worry about how to render "facts" about our "mental" lives consistent with our physical understanding of the rest of Nature, we should instead appreciate that there are no such facts to begin with. At the other end of the spectrum is the "New Mysterianism" which claims that due to contingent facts about our cognitive capacities it is conceptually impossible for us to understand how mental states are physically realized, even though they are physical phenomena.[6]

The philosophical terrain of physicalism is thus variegated indeed, but despite the fact that these theories all have different approaches to the mind and therefore have different conceptions of how it is physically realized, they all agree with the basic premise that physicalism is true. The view that human beings are composed of completely physical substance and that their behaviour is explainable in terms of physical concepts has gained wide acceptance and (as the range of different forms of physicalism attests) has become much more complex than the early materialism of Descartes' critics. This is in no small part

---

[5] Paul Churchland, Matter and Consciousness (Cambridge: MIT Press, 1984); William Ramsey, Stephen Stich, and Joseph Garon, "Connectionism, Eliminativism, and the Future of Folk Psychology," in The Future of Folk Psychology: Intentionality and Cognitive Science, ed. J. Greenwood (Cambridge: Cambridge University Press, 1991).

[6] Colin McGinn, The Problem of Consciousness (Cambridge: Basil Blackwell, 1991).

due to changes in what science countenances as physical phenomena. The world of contemporary physics, with its fields of force and quantum indeterminacies, is a long way from the material universe described by Newton or envisioned by LaPlace. This is why most philosophers today speak of "physicalism" rather than "materialism." For physics now allows for items in its ontology that are not straight-forwardly material, such as electromagnetic fields and the like. The advantage of moving away from a mechanistic billiard-ball model of the universe is that it gives science more flexibility and resources to explain phenomena, especially phenomena as complex as human thought and experience.

With advancements in physics, and especially neurology, there is mounting evidence to suggest that the mind can be understood scientifically and there are even clues as to how the mind is physically realized. Since the work of Pierre Paul Broca[7] on aphasia (a disorder involving the use or comprehension of language due to cerebral damage often resulting from stroke) there has been much progress in localizing higher cognitive functions in specific areas of the brain. Broca concluded that the left hemisphere is responsible for speech, and subsequent work by Gustav Theodor Fritsch and Eduard Hitzig,[8] and later Carl Wernicke,[9] offered more precise localizations of different elements of language comprehension and use, such as word selection. Also, research on temporal lobe epilepsy has provided strong evidence in favour of the idea that certain parts of the brain are responsible for specific types of emotion. For example, patients with right temporal lobe epilepsy tend to be excessively emotional, whereas those with left temporal lobe epilepsy tend to "manifest ideational traits such as a sense of personal destiny, moral self-

---

[7] Pierre Paul Broca, "Sur le siége de la Faculté du langage articulé," Bull. Soc. Anthropol. 6 (1865) pp. 377-393.

[8] G. Fritsch and E. Hitzig, (1870) "Ueber die elektrische Erregbarkeit des Grosshirns," in Some Papers on the Cerebral Cortex, ed. G. von Bonin (Springfield, Ill.: Thomas, 1960).

[9] Carl Wernicke, (1908) "The Symptom-Complex of Aphasia," in Diseases of the Nervous System, ed. A. Church (New York: Appleton).

scrutinizing, and a penchant for philosophical explanation."[10] The evidence of such localizations is often used as anecdotal support for particular versions of physicalism.

Despite the advancements in neurology, however, there remain a number of stubborn philosophical problems that appear to resist solution in the effort to fill in some of the details of a physicalist account of mind. The philosophers who take these problems seriously offer compelling reasons to think that these difficulties introduce a fundamental challenge to physicalism. My aim in this dissertation is to explore and to evaluate these difficulties.

Contemporary arguments against physicalism attempt to show that forms of physicalism are incomplete because there are facts about human beings that cannot be captured or explained in physical terms. The arguments I examine attack physicalism on two fronts, corresponding to the distinction commonly recognized between two sorts of mental states or events: the propositional attitudes and qualitative states of consciousness (qualia). Propositional attitudes are mental states that possess propositional content such as beliefs, desires, fears, thoughts, doubts, and so on. The reason they are called "propositional attitudes" is that they have propositions, or statements, as their intentional objects. If I fear or doubt something (the "attitude") there must be some propositional object to which my fear or doubt is directed; that is, something that I fear or doubt (e.g., "that I have forgotten my appointment," or "that it will be a mild winter"). Qualitative states of consciousness, on the other hand, are often thought not to possess propositional content though such states can be identified propositionally. For instance, I can refer to my visual experience of a red object by saying that I see something red, but there is more going on in my conscious awareness than the mere entertaining of a proposition: there is also the

---

[10] Eric R. Kandel, "Brain and Behaviour," in Principles of Neural Science (2nd edition), ed. Eric R. Kandel and James H. Schwartz (New York: Elsevier Science Publishing Co., Inc., 1985) p. 10.

intrinsic character of the experience itself—how the red object looks or appears to me on that occasion. By bringing together arguments that challenge the ability of physicalism to handle both types of mental phenomena I examine a thorough and comprehensive critique of physicalism.

The attacks on physicalism that focus on the propositional attitudes are typically connected with issues concerning mental causation. If physicalism is true, then the reasons, desires, wishes, and other mental states that we ordinarily take to cause and to explain our intentional behaviour are physical events and have the power to cause our behaviour in virtue of their physical composition and properties. But if this is the case, then it seems as though the fact that our psychological states have the propositional contents they do has little or nothing to do with the causes of our behaviour. For it is not in virtue of my desire being the particular desire it is, or having the propositional content it does that causes me to act in a certain way. Rather, the causal efficacy of my desire is owed entirely to its neurological properties. Thus, despite the fact that we ordinarily think that our psychological states cause our behaviour it seems to follow from some forms of physicalism that they have no causal powers at all. The view that our mental states or certain of their properties (their propositional contents) do not cause behaviour is often called "epiphenomenalism." If reasons and the like are epiphenomenal (i.e., lacking causal powers) then this creates a difficulty for physicalism because there is a great deal of evidence to suggest that our mental states do cause our behaviour. Most of the time an explanation of someone's actions in terms of his or her reasons is the best explanation that we have. It is therefore difficult to see how physicalism can maintain both that our behaviour is caused by purely physical features of the brain and that we can explain an action by identifying the propositional content of someone's beliefs and desires. It begins to appear as though there are aspects of human behaviour that escape capture in physical

terms, for it seems that certain kinds of explanation function quite independently of physical explanations.

The difficulties with the qualitative states of consciousness represent a similar obstacle to physicalism. There appears to be no way to capture in physical terms the intrinsic feelings of the experiences with which we are directly acquainted, for attempts to account for the physical nature of experience always seem to leave something out of the story. For example, while a neurologist might be able to tell us all about the physical processes involved in colour vision, it seems that one cannot thereby learn about the intrinsic character of the experience of a certain shade of red. Once again, then, it seems that there are features of mentality that somehow escape physical capture, which suggests that such features are epiphemenomenal. For to deny that the qualitative states of consciousness can be discovered by studying the causal mechanisms of perception is apparently to deny that they play any causal role in colour perception. Furthermore, given that there are facts about human experience that cannot be captured physically it seems to follow that there are non-physical facts, and hence that physicalism is false.

My aim in what follows is to examine the details of the epiphenomenalist arguments and to suggest that they all move illegitimately from the claim that physicalism suffers from various explanatory inadequacies to the conclusion that physicalism is false. I will stipulate that physicalism as a general thesis consists of two central claims. First, there is the ontological claim that whatever exists is ontologically physical and that human beings are among such physical things. Second, there is the further claim that physicalism is explanatorily complete: that all facts about human beings and all explanations of their behaviour are dependent on and determined by physical facts and explanations. While the ontological claim is uncontroversial as one constituting physicalism, the second is not. It is therefore advisable to say something about my reasons for insisting on it. My formulation of physicalism is motivated by two considerations. First and most importantly, the forms

of physicalism that have been subject to the epiphenomenalist challenges accept both these claims, either explicitly or implicitly, and since my task is to evaluate these objections it only makes sense to work with the understanding of physicalism under discussion. Second, the commitment to explanatory completeness seems a necessary requirement for a robust version of physicalism. I grant that weaker articulations lacking this second condition are possible and that there might be good reasons to endorse these weaker versions of physicalism, but if physicalism is to be a philosophically interesting thesis, then it seems it should involve more than the bald ontological claim; it should include the idea that our explanations of phenomena (whether they make use of physical predicates or not) are dependent on and determined by physical states of affairs. Thus, my intention is that this understanding of physicalism fit with the idea, shared among most physicalists, that physical facts determine all the facts.

The arguments against physicalism I consider are primarily directed against (2). However, there is one exception to this (Frank Jackson's argument, discussed in Chapter Three), but I show that the ontological conclusions of the argument in fact do not follow; the strongest plausible conclusion one can draw from the argument is the denial of (2). The next two chapters of this dissertation are therefore devoted to making as strong a case as seems possible against the explanatory completeness of physicalism. While I claim that the arguments against (2) are initially compelling, I show in the remainder of the dissertation that they are undermined if physicalism can incorporate a properly formulated thesis of psycho-physical supervenience. In the end, it should be clear that the epiphenomenalist arguments against physicalism are without any force at all.

I will now sketch out the framework of my discussion in more detail. Since the various epiphenomenalist challenges are not raised against any one particular version of physicalism, I survey several different articulations of the theory and the points where the criticisms arise. Historically one can trace the development of physicalism as a thesis which

has been gradually weakened over the course of time, and my investigation of the challenges to physicalism will reflect the course of this development.

In Chapter Two I examine Donald Davidson's anomalous monism, a well-known and extremely controversial version of the identity theory. Davidson claims that mental events such as thoughts, perceptions, reasons and desires are token-identical to physical events. That is, each particular mental event is identical to some particular physical event. There is, in his view, nothing intrinsically mental about mental events; they are simply physical events characterized under mental descriptions. Many of Davidson's critics have argued that his account of the identity relation between mental and physical events, together with his view of causation, entail that mental properties play no causal role in the production of behaviour. As I mentioned earlier, the worry is that given Davidson's physicalism it appears as though all the causal work is performed by the physical properties of one's mental states, in which case the fact that one's beliefs, desires and intentions were the particular beliefs, desires and intentions they were begins to seem irrelevant to one's actions. Hence, one's reasons and other mental states are epiphenomenal. The fact that the mental events in question are identical to physical events does not render the mentioned mental properties causally efficacious. For we can say that the neural event alone—or the physical properties that constitute it—would have been sufficient for the behaviour in question, and therefore the fact that one had a reason for doing what one did contributed nothing to the action.[11] In any case, we cannot say that the action was caused just in virtue of the fact that the neurological state had the propositional content or mental properties that it did.[12] So appeals to identity are no help in alleviating this problem.

---

[11] See Stephen Yablo, "Mental Causation," Philosophical Review 101 (1992) pp. 245-280.

[12] Frederick Stoutland, "Oblique Causation and Reasons for Action," Synthese 43 (1980) pp. 351-367.

Many of Davidson's critics take this to be a decisive blow against anomalous monism as an articulation of physicalism. In the first place, it shows that Davidson has not successfully grounded the psychological explanations of behaviour we ordinarily appeal to in a physicalist scheme. Also, if Davidson's critics are correct, then mental properties are cut off from the physical world of causes and effects since they have no effects in the physical world. Such properties are surely very strange and difficult to reconcile with a physicalist ontology. For the world of physics is the world of causes and effects. Finally, this account of the production of intentional behaviour seems very far from what most of us would regard as an adequate account of mental causation, for the mental seems to drop out of the story altogether.

In the first section of Chapter Two I examine this objection in detail and show that there is good reason to think it is completely misconceived. I argue that the epiphenomenalist objection as it is usually formulated presupposes a view of events and of causation that Davidson does not endorse. As I have already indicated, the standard objection that anomalous monism entails epiphenomenalism of the mental builds on the idea that the causal efficacy of an event is owed solely to its physical properties, from which it allegedly follows that mental properties have no causal powers. This idea requires that one think of events as being constituted by clusters of properties, some of which have causal influence and others of which don't. If one accepts such a view of events and of causal relations the epiphenomenalist conclusion is difficult to resist. However, there are indications that Davidson does not subscribe to this view, nor is he required to for reasons of internal consistency. I argue that Davidson is an anti-realist about mental properties (and tries to be one about physical properties too) in which case the standard criticism of Davidson's view doesn't apply. I conclude that anomalous monism does not in fact entail epiphenomenalism in the traditional sense.

Despite my defense of Davidson from the standard objection to anomalous monism I think there is a genuine problem behind the criticism. In the second section of Chapter Two I reformulate the epiphenomenalist objection in a way that appears to have some damaging consequences for Davidson's theory while doing justice to the main intuitions of the standard objection. To some extent this requires reworking the very notion of epiphenomenalism. I argue that Davidson's theory does not entail that mental events or properties are epiphenomenal in the sense of lacking causal efficacy, but in a slightly different sense. This alternative understanding of epiphenomenalism is that although mental events cause behaviour they do not causally explain behaviour. Hence, mental events (or properties) have no role to play in the causal explanations of intentional action. By examining Davidson's account of explanation and causation I show that in light of his anomalism of the mental and some general constraints on the nature of explanation, Davidson's theory entails this alternative version of epiphenomenalism. This form of epiphenomenalism appears to have consequences for the explanatory completeness of Davidson's brand of physicalism. The worry is that the denial that reason-giving is a species of causal explanation suggests that psychological explanations are autonomous since they cannot be connected to causal explanations dealing in physical predicates. The apparent autonomy of such explanations entails the denial of physicalism's claim to explanatory completeness since there are explanations that appear not to depend on physical explanations.

In Chapter Three I examine the epiphenomenalist challenges centering on the qualitative states of consciousness. The first section involves an exploration of a challenge that is frequently made to the theory of mind known as "functionalism." Functionalism is a form of identity theory which claims that mental states are identical to functional states of complex systems which can be realized by a variety of different physical bases. Mental states are therefore abstract organizational states rather than specific neurological states

which can in principle be instantiated by computers, silicon brains, or other forms of life. The most infamous challenge to functionalism is the "inverted spectrum problem." The inverted spectrum hypothesis, which has its origins in the philosophy of John Locke, suggests that it is possible for two systems to be functionally identical yet differ in some mental respect. For example, consider two individuals who respond to the colour red in the same way; they make all the same colour discriminations and consistently describe the relations between red and other colours. Despite the similarity in behaviour of our two individuals, it is possible to imagine that one person might nevertheless have a completely different subjective experience of colour from the other. While one person sees red things as red, the person experiencing the inverted spectrum has the sort of experience ordinarily associated with green objects. Because there are no behavioural cues to betray the difference in the phenomenal character of the mentioned experiences the functionalist must say these two people are in the same mental state even though they are not. This suggests that functionalism is unable to accommodate the qualitative characteristics of subjective experience, and hence that it is incomplete in its attempt to capture and characterize the mental.

A number of philosophers have argued that spectrum inversion is impossible, but such arguments typically require the acceptance of questionable assumptions about the semantics of colour words or work on excessively verificationist grounds. I propose a less ambitious and less questionable approach to this idea. I argue that the idea of an undetectable spectrum inversion does not constitute a plausible hypothesis, in which case it offers no reason to think that the qualitative states of consciousness represent a fundamental challenge to functionalism. My strategy is to show that sensory states have affective contents which are partly constitutive of sensations. Given this, if one's qualia are inverted, then it follows that the affective content of the sensation will be inverted also, since they are not radically separable. Because psychologists have demonstrated great ingenuity in

designing tests to measure emotional responses to stimuli, I suggest that there is good reason to expect that the inversion of emotional state that accompanies colour qualia inversion will manifest itself in behaviour, and hence, that the underlying phenomenal difference can be functionally captured. After a brief discussion of colour vision and the affective states that are connected with the experience of certain colours, I examine the other senses and show that there are similar reasons to believe that the qualia connected with them are also partly constituted by emotional content, and that such constitution rules out the likelihood of undetectable qualia inversions for those senses as well. Thus, we shall see that there is little reason to suppose that an undetectable inversion of phenomenal elements represents a genuine possibility for any sensory modality. The failure of the inverted qualia hypothesis therefore salvages functionalism and suggests that qualia represent no special obstacle to physicalism.

In the second half of Chapter Three I explore another qualia-based objection to physicalism. This time the objection is Frank Jackson's "knowledge argument."[13] Jackson's argument is more interesting than the inverted spectrum hypothesis because of the breadth of the conclusion. The knowledge argument is not geared toward any one particular version of physicalism, but against physicalism generally. Jackson claims to show that there are facts about human experience (again, facts about the qualitative states of consciousness) which cannot be captured in physical terms, and hence that there are non-physical facts. Furthermore, Jackson argues that the non-physical properties to which these facts refer are epiphenomenal in the traditional sense of being effects of physical states but themselves being causally impotent. He arrives at this conclusion by means of a thought-experiment. He asks us to imagine that a brilliant neurologist, Mary, who despite living her

---

[13] Frank Jackson, "Epiphenomenal Qualia," Philosophical Quarterly, 32 (1982) pp. 127-136.

entire life in a black and white environment comes to learn everything physical there is to know about the neurophysiology of colour vision. Thus Mary knows all the physical information there is to know about seeing red, for example. When Mary leaves her room for the first time Jackson claims that she will learn something about the experiences of others that she did not know before. She will learn what it is like to see the colour red. Since Mary knew all the physical information about colour vision before her escape yet comes to possess new information afterwards, it follows that she learns a non-physical fact. This entails that there are non-physical properties and that such properties are epiphenomenal, for otherwise Mary would have noticed mysterious gaps in her previous understanding of the causal processes involved in colour vision.

While Jackson's argument is intuitively plausible, and most of the standard replies to Jackson are unconvincing, I show there are reasons to soften the conclusion that can legitimately be drawn from the knowledge argument. I demonstrate that it is possible for properties like phenomenal redness to escape capture in causal accounts of colour vision without being led to the conclusion that such properties are non-physical. The knowledge argument therefore does not challenge the ontological thesis of physicalism as Jackson thinks. However, it remains true that the physicalist cannot capture the intrinsic character of certain experiences, such as seeing red. It therefore seems that Jackson's argument entails the weaker conclusion that physicalism is incomplete at the explanatory level. The conclusion drawn in Chapters Two and Three is therefore the same: the epiphenomenalist objections to physicalism appear to call the explanatory completeness of physicalism into question. In Chapter Four I invoke the concept of supervenience to show that in fact this conclusion does not follow. The identified incompleteness of physicalism is not substantive, but merely apparent.

In Chapter Fourt I examine the weakest and most recent formulation of physicalism: supervenience. While psycho-physical supervenience goes back at least as far as

Davidson's token-identity theory, there has recently been a proliferation of different forms of supervenience in the philosophical literature. Supervenience expresses a relation that holds between sets of predicates or properties. Typically the relation is thought to be one of dependence and determination, such that one's mental properties are determined by and dependent on one's physical properties. If supervenience can be shown to include these ideas, then it is a very useful tool that can be employed to address the problems identified in the earlier chapters.

I discuss two central aspects of supervenience, both of which are very important if supervenience is to represent a form of physicalism that is strong enough to avoid epiphenomenalism. The first aspect of supervenience I consider is the assumption made by many that this relation (in the philosophy of mind) expresses the dependence of the mental on the physical. This is generally thought to be the weakest possible form of physicalism since it does not necessarily entail an identity or reduction of mental properties to physical properties. Nevertheless, the relation of dependence gives the physical ontological priority, which for most philosophers is enough to avoid dualism or epiphenomenalism. However, Jaegwon Kim has raised a number of powerful arguments against the claim that existing formulations of supervenience express dependence at all. If his arguments are compelling, then they represent a serious blow to the possibility of conceiving of physicalism in terms of supervenience.

In my evaluation of Kim's arguments I examine several different characterizations of psycho-physical supervenience: those developed by Kim himself, and Donald Davidson's. I show that despite some confusions Kim is essentially correct that the alternative formulations of supervenience he considers do not express psycho-physical dependence. Davidson's model of supervenience, however, is more promising than Kim's in this regard. Building on some themes from Chapter Two I show that Davidson's conception of supervenience is better thought of as a semantic thesis connecting predicates

(as opposed to Kim's, which is a metaphysical thesis about properties), which, unlike Kim's, does give us a relation of dependence. However, the kind of dependence expressed is not the metaphysical dependence most philosophers are looking for; instead it is a kind of semantic dependence that can be derived from Davidson's treatment of mental ascription.

In the second section I show how the Davidsonian understanding of supervenience just discussed can be employed to undermine the force of the epiphenomenalist objections developed in the earlier chapters. In brief, the approach is as follows. It has been suggested that reason explanations and facts about qualia are autonomous, i.e., they do not depend on physical explanations or facts, and it is because of this that physicalism is incomplete at the explanatory level. However, qualia and reason explanations supervene on physical facts and explanations. Supervenience is a relation of dependence, therefore qualia and reason explanations depend on physical facts and explanations after all. It follows, then, that physicalism is complete at the explanatory level after all.

In schematic form, then, the line of argument in this dissertation runs roughly as follows:

1. Physicalism (the ontological claim plus the claim to explanatory completeness) is true.
2. There are facts and explanations that do not depend on physical explanations.
3. Therefore, physicalism is not explanatorily complete.
4. Hence, (1) is false.
5. But if we accept the Davidsonian-style thesis of supervenience, then (2) is false.
6. Therefore, (3) and (4) do not follow.

In Chapter Five I briefly take up the implications of these conclusions for work in the philosophy of mind and point to some further issues which will have an important bearing on the topics discussed in this dissertation. In particular, I introduce some possible difficulties for anomalous monism which question Davidson's ability to handle the objections I have mentioned in a consistent manner. Although my suggestions here will fall

short of constituting an adequate physicalist theory of mind they will nevertheless point to some useful questions and possible answers, which, if pursued should lead to a viable form of physicalism.

Chapter 2

Anomalous Monism and Epiphenomenalism

The first form of physicalism which has been the object of epiphenomenalist worries that I want to consider is Donald Davidson's anomalous monism. Davidson's view can be seen as emerging from two pressures on its predecessor, the type-identity theory. The type-identity theory claimed that mental types such as pain are identical to types of physical states. This involved a theoretic identification of types along the lines frequently seen in the sciences as, for example, when lightning is said to be identical to a rapid discharge of electrons in the atmosphere. When scientists make such an identification they are not saying that there are two things that are correlated, they are saying that lightning is nothing more than a rapid discharge of electrons in the atmosphere, or that lightning is reduced to a rapid discharge of electrons in the atmosphere. Similarly, in the case of the type-identity theory the idea was that mental types could be identified with, and thereby reduced to, physical types such as kinds of neurological or neurophysiological states.

The pressures on the identity theory I referred to cluster around this notion of the reduction of mental types to physical types. I will not here evaluate the cogency of these arguments since my intention is simply to provide some background for my discussion of anomalous monism. The first of these pressures comes from the fact that the proposed reduction of mental types to the physical types presupposes a model of science which has, since its hey-day in the late 1950s, fallen out of favour. This model is known as "The Unity of Science." The central claim constituting the unity of science was that the sciences could be unified, via a chain of reductions, in the sense that all the sciences could, through

a series of steps, be reduced to the most basic science (physics). For instance, the idea was that sociology could be reduced to psychology, psychology to neurology, neurology to biology, biology to chemistry, and chemistry to physics. Since the reduction was thought to be transitive it was believed that through a chain of reductions such as the one just mentioned, higher-level sciences such as sociology or psychology could be reduced to physics. The reductions themselves involved demonstrating that the predicates of one theory are co-extensive with the predicates of a more basic scientific theory, in much the same way that lightning was shown to be co-extensive with a rapid discharge of electrons in the atmosphere. Once these "point reductions" were established the aim was to model the laws of the theory to be reduced to laws of the more basic reducing theory, which would render the reduction of one science to the other complete.[1]

Unfortunately the optimism with which people first embraced the idea of the unity of science has proven to be ill-advised. Although there have been a number of significant developments in the sciences since the 1950s there has in fact been little movement toward the unity of science. Instead, there appears to be more of a tendency to regard the various levels and branches of science as autonomous. Some, such as Ian Hacking, have suggested that given this the unity of science does not reflect scientific practice at all but is instead merely a philosopher's "idle pipedream."[2] Since scientific practice does not appear to support the principle of the unity of science, there is little reason to expect the reductions promised by the type-identity theory.[3]

---

[1] See Oppenheim and Putnam, "Unity of Science as a Working Hypothesis," in The Philosophy of Science, ed. Richard Boyd, P. Gasper, and J.D. Trout (Cambridge: MIT Press, 1991).

[2] Ian Hacking, "Weapons Research and the Form of Scientific Knowledge," in Nuclear Weapons, Deterrence and Disarmament, ed. D. Copp (Calgary: University of Calgary Press, 1986).

[3] See William Seager, Metaphysics of Consciousness (London: Routledge, 1991).

The other pressure on the type-identity theory is related to the one just mentioned but is more philosophical in content. If we assume that the identity theory is true, then this means that mental types such as pains are certain neurological states. This is typically expressed in the form of a biconditional: X is in pain if and only if X is in neurological state Y. This means that anything in neurological state Y is in pain, and anything in pain is in neurological state Y such that it is impossible for something to be in pain yet not be in the specified neurological state. Such a claim has been criticized as representing a kind of chauvinism since it precludes the possibility of other life-forms which are physically different from us from having psychological states like ours.[4] Since octopi and (if there are such things) Martians do not have brains like ours they cannot be in neurological state Y, which means they cannot feel pain. Nevertheless, the critics claim it is highly implausible to deny that such forms of life can experience pain, in which case pain cannot be identical to a particular neurological state unique to humans. This is referred to as "the problem of multiple realization." The idea is that if we are going to identify mental states with physical states we need to allow for the possibility that creatures with a variety of different physical structures and compositions can share our mental states. Therefore, pain must be realizable in a multitude of different physical states, which gives us another reason to doubt the claim that mental types can be reduced to physical types.[5]

Davidson's anomalous monism emerges as a form of identity theory which is sensitive to these pressures, for the main feature of his theory is that he claims mental events are identical to physical events, yet denies that the mental is reducible to the physical. The way Davidson achieves this is by maintaining that the identity holds between

---

[4] See Ned Block, (1978). "Troubles with Functionalism," in Readings in Philosophy of Psychology Vol. 1, ed. Ned Block (Cambridge: Harvard University Press, 1980).

[5] For a more formal discussion of this point see Jerry Fodor, "Special Sciences," in Boyd et. al.

mental and physical tokens (particular mental and physical occurrences) instead of types. So, according to Davidson, when I experience a pain that pain is identical to some particular physical state in me, and when you are in pain that particular pain is identical with some physical state in you, but this does not mean that when we are both in pain there necessarily exists some physical state that we share. Pain is physically realized in a different way in me than it is in you, and it is even possible for that physical realization in each of us to change over the course of time, so it is not even true that when I am in pain on two occasions that I am necessarily in the same physical state on both occasions. Since Davidson denies that an identification can be made between mental and physical types, he denies that mental concepts, such as pains, can be reduced to physical concepts. Davidson's theory, then, represents a form of nonreductive materialism.

Davidson's account of the relationship between mental and physical events has given rise to a small industry of criticism. The theme common to most of this work is the suggestion that anomalous monism entails epiphenomenalism. My aim in this chapter is to explore this charge. The discussion of this topic is divided into two parts. In the first section I sketch out the standard argument against Davidson's anomalous monism and show why it is unconvincing. In the second section I reformulate the argument in a way that avoids the shortcomings of the standard version of the objection and draw out the consequences this criticism seems to have for Davidson's physicalism.

1. The Standard Objection to Anomalous Monism[6]

Since he first proposed it in "Mental Events"[7] numerous authors have criticized Davidson's account of the relation between mental and physical events. The usual charge against Davidson is that anomalous monism renders mental properties "epiphenomenal" or "causally inert." The standard form of this objection claims that Davidson's account of causation entails that it is only in virtue of the properties picked out in a physical description that events can instantiate strict causal laws. Since physical properties are the only ones that figure in strict laws, it follows that they are the properties in virtue of which events cause; therefore, even if we assume that mental events are token-identical to physical events, the mental properties of an event contribute nothing to the causal efficacy of that event and are consequently epiphenomenal. In his recent article "Thinking Causes"[8] Davidson at last offers a response to his critics on this point. In his view the epiphenomenalist objections depend on a misunderstanding of his account of events and causation. So-called properties should not, as his critics assume, be thought of as things in the world. Instead, one should really speak of predicates, and what predicates are ascribed to an event is a matter of how the event is described. (This should not be interpreted to mean that properties just are predicates or descriptions. For Davidson properties aren't anything. The point is rather that when people speak of properties in the sense that Davidson's critics do, from Davidson's

---

[6] A version of this section is to appear under the same title in Australasian Journal of Philosophy 75 (1997).

[7] Donald Davidson, "Mental Events" (1970) in Davidson Essays on Actions and Events (Oxford: Clarendon Press, 1980).

[8] Donald Davidson, "Thinking Causes" in Mental Causation, ed. John Heil and Alfred Mele (Oxford: Clarendon Press, 1995).

point of view they ought really to talk about predicates.) The standard objection turns out to be misguided, then, because for Davidson it doesn't make sense to say events stand in causal connection in virtue of certain properties.[9] For on Davidson's view, since "causality is a relation between events, it holds no matter how they are described."[10] Davidson's critics remain unconvinced by this reply and continue to urge that his nomological account of causation requires talk of properties and not merely of descriptions.

My aim in this section is to clarify the source and terms of this debate and to offer steps toward its resolution. This needs doing because the responses to Davidson's recent defense of anomalous monism indicate that Davidson and his critics are farther apart than ever on the question concerning the causal efficacy of the mental. I think the key to making progress in this dispute lies in properly recognizing Davidson's belief that events are not constituted by clusters of recognition-transcendent properties. On Davidson's view there is nothing "in" events that explains why events support certain descriptions as opposed to others. This claim is consistent with Davidson's adoption of a Tarski-style semantics of truth. Davidson says, "Nothing,...no thing, makes sentences and theories true: not experience, not surface irritations, not the world, can make a sentence true."[11] We can generalize this claim and say also that "Nothing, no thing makes it true that an event supports certain descriptions or can be described using certain predicates." This stems from Davidson's refusal to reify properties. On Davidson's account properties are not ontological constituents of events which determine the truth or falsity of our descriptions. There are, then, no properties in virtue of which events support certain descriptions, and hence, no properties in virtue of which events cause.

---

[9] Ibid., p. 13.

[10] Ibid., p. 6.

[11] Donald Davidson, "On the Very Idea of a Conceptual Scheme," in Davidson Inquiries into Truth and Interpretation, (Oxford: Clarendon Press, 1984) p. 194.

Provided we understand the term in a sufficiently broad sense, I would say that this conception of events involves an anti-realist attitude toward properties. I use the term "anti-realist" to emphasize Davidson's denial of the recognition-transcendence of properties. In Davidson's view there are no facts about what properties events have independently of us describing events in a certain way. There is a sense, then, in which the properties (or better, predicates) possessed by an event (or true of that event, as described) depend upon the event being recognized as having such properties.[12] What needs to be appreciated by Davidson's critics is that this anti-realist attitude toward properties underlies Davidson's response to the property-based epiphenomenalist objections. Although the claim that events are not constituted by realist properties would completely dissolve the argument against anomalous monism, I show that Davidson's other commitments make this idea problematic, and that his critics actually have a point in their favour. However, I also show that Davidson's anti-realism about the mental is sufficient to block the epiphenomenalist objection in its usual formulation.

Before I offer my suggestions about the source of the disagreement over anomalous monism and its possible resolution, it would be helpful to outline Davidson's view and the standard objection in more detail. The thesis of anomalous monism states

> that mental entities (particular time- and space-bound objects and events) are physical entities, but that mental concepts are not reducible by definition or natural law to physical concepts.[13]

---

[12] While I think this idea makes it plausible to speak of Davidson as an "anti-realist" toward properties in this loose sense (especially in connection with mental properties, as we shall see) it should be acknowledged that more work would have to be done to show that Davidson's refusal to reify properties thereby makes him an anti-realist in the more technical sense. For, one might suspect that the relegation of properties to recognition-dependent entities turns properties into things, and as I already suggested one should not interpret Davidson as accepting this. Sorting out these issues goes well beyond the scope of this section but would certainly be worth while exploring.

[13] Davidson, "Thinking Causes," p. 3.

The key to this idea is well known. The identity between mental and physical events is a token-identity. Since the identity holds between tokens instead of types there are no psycho-physical bridge-laws which would support the reduction of mental concepts to physical concepts. Davidson derives the theory of anomalous monism from three premises:

> (1) that mental events are causally related to physical events, (2) that singular causal relations are backed by strict laws, and (3) that there are no strict psycho-physical laws.[14]

Davidson's critics have long argued that these premises are inconsistent. The objection is that (2) and (3) together exclude the truth of (1).

According to most critics the source of the inconsistency is ultimately premise (2). This premise indicates Davidson's espousal of Hume's nomological account of causation. In "Mental Events" Davidson encourages a weak reading of Hume's claim that a causal law "covers" every singular causal claim. According to Davidson by this we should not take Hume to mean that the statement of the relevant covering law is necessarily formulated in the same terms as the singular causal claim; instead, we should take him to mean that the statement of the law incorporates some true description of the events related as cause and effect. For Davidson this would be the description provided by the "closed system" of an ideal physics. An ideal physics constitutes a closed system because the descriptions of events possible in that language are fully extensional and express exceptionless laws free from intrusion by intensional concepts. In the case of singular causal claims involving mental events—claims fitting premise (1)—the relevant covering law cannot be formulated in the same terms as the singular causal claim. The reason for this is that for Davidson our mental concepts do not constitute a closed system in the way those of a complete physics would, and hence are not amenable to the formulation of strict laws. The source of this belief lies in Davidson's interpretationalism. Since our ascription of mental states and

---

[14] Ibid.

events to agents is always open to reinterpretation over the course of time, mental events cause behaviour only as mediated by other mental events (namely, those ascribed to an agent at a later time) "without limit."[15] Because strict laws, being exceptionless, require fixed and determinate descriptions, the unruly and indeterminate behaviour of mental descriptions excludes their participation in strict laws. This means that the law covering causal claims involving mental events must be formulated in physical, not psychological terms.

By denying that the mental constitutes a closed system Davidson makes a number of questionable assumptions about the character of physics. For example, he assumes that physical measurements (which clearly belong to the physical scheme) are determinate and repeatable. On this view, were we to discover some discrepancy in our measurements we would think this happened not because the physical object being measured is indeterminate (in the way mental states are) with respect to the characteristic we are interested in; instead, such discrepancies would be attributed to errors in observation or the improper functioning of our instruments. Even in the case of the radical changes in measurement possible with scientific revolutions the assumption is that the physical facts themselves don't change, it is our way of capturing them that does. While these assumptions are questionable and we might wonder whether there really is a significant distinction between the mental and physical schemes, I will grant Davidson this assumption since in objecting to him we ought to share as many assumptions as possible. I will therefore assume that Davidson is correct in his claim that the law covering causal claims involving mental events must be formulated in physical terms.

Davidson's critics take this to show that the mental properties of events have no real causal efficacy. Their reasoning is that since it is only under a physical description that an

---

[15] Davidson, "Mental Events," p. 217.

event can instantiate a causal law it is only as a physical event that a mental event can cause anything. In other words, the nomological character of causality commits Davidson to the view that mental events cause only in virtue of the properties picked out by a physical description and not in virtue of their mental properties; therefore, mental properties are causally inert.[16] This represents a problem because this account of how mental events cause seems very far from what we would ordinarily take to be an instance of mental causation. On Davidson's account it would appear as though the fact that my mental event was the particular mental event it was or had the particular mental properties it did (say, a desire for a cold drink) contributes nothing by way of causing me to go to the refrigerator. Instead, it is the physical properties underlying my desire for a cold drink that do all the causal work.

Davidson's response to this objection is that it misrepresents his position. As he puts it,

> . . . if causal relations and causal powers inhere in particular events and objects, then the way those events and objects are described, and the properties we happen to employ to pick them out or characterize them, cannot affect what they cause. Naming the American invasion of Panama "Operation Just Cause" does not alter the consequences of the event.[17]

Unfortunately, by speaking of "the properties we happen to employ to pick them out or characterize them," Davidson invites the very confusion surrounding the status of properties that has fed this debate for so long. It is therefore desirable to reiterate that in Davidson's considered view, talk of "properties" is really just talk about predicates and what predicates are true of an event is a matter of how the event is described. There is

---

[16] For example see Peter Hess, "Actions, Reasons, and Humean Causes," Analysis 41 (1981) pp. 77-81; Ted Honderich, "The Argument for Anomalous Monism," Analysis 42 (1982) pp. 59-64; Terence Horgan, "Mental Quausation," Philosophical Perspectives 3 (1989) pp. 47-76; Jaegwon Kim, "Can Supervenience Save Anomalous Monism?" in Mental Causation, ed. John Heil and Alfred Mele (Oxford: Clarendon Press, 1995).

[17] Davidson, "Thinking Causes," p. 8.

therefore no question about the causal significance of such items. Whether we use one description or set of predicates as opposed to another when we identify an event has no bearing on that event's causal powers. The epiphenomenalist objection, then, is misguided in its attempt to explain causation in terms of the efficacy of properties because events, not descriptions, cause.

Davidson's critics, most notably Jaegwon Kim, remain unconvinced by this reply. Despite Davidson's rejoinder, Kim continues to insist on an account of the causal efficacy of events in terms of the efficacy of properties:

> The issue has always been the causal efficacy of properties of events—no matter how they, the events or the properties, are described. What the critics have argued is perfectly consistent with causation itself being a two-termed extensional relation over concrete events; their point is that such a relation isn't enough: we also need a way of talking about the causal role of properties, the role of properties of events in generating, or grounding, these two-termed causal relations between concrete events.[18]

Since Kim thinks Davidson must recognize that the causal efficacy of events is derived from the properties of events, and that consequently only certain properties will have causal relevance, the epiphenomenalist objection stands.

Davidson and his critics, then, are divided on the question of whether one should account for the causal connection between events in terms of events or properties of events. Davidson thinks talk of properties of events is superfluous and confusing, and his critics think causal relations are mysterious without the possibility of such an account. The source of this difference is that Davidson thinks events cannot be broken down into recognition-transcendent properties. For Davidson's espousal of a Tarski-style semantics of truth implies that there is nothing in virtue of which events have certain properties or support certain descriptions. Kim, on the other hand, holds the view that an event is constituted by

---

[18] Jaegwon Kim, "Can Supervenience Save Anomalous Monism?" in Heil and Mele, p. 21.

three components: an object, a property, and a time.[19] Given this characterization of events it would seem only natural for Kim to insist on an analysis of causal relations in terms of the efficacy of properties. For since events are constituted in part by properties, the causal powers of events can be explained in terms of the causal powers of the properties that constitute them. Since this would allow us to distinguish those properties of an event that are causally relevant from those that are not, we can describe the latter as epiphenomenal. However, since Davidson does not accept this characterization of events, Kim's objection really begs the question. To make a convincing case against Davidson Kim must show either (1) that Davidson ought to conceive of properties and events in the same way he does, or (2) that his objection goes through even on Davidson's account of properties and events. (2) does not have much prospect for success, since on Davidson's conception of events it "makes no literal sense"[20] to speak of properties in virtue of which events stand in causal relation; for on this view of events there are no such properties.

It seems that the only way to make the objection work is to adopt the first strategy and show that Davidson ought to treat properties in the way Kim does, as real ontological components of events. One way of doing this would be to demonstrate that Davidson's other commitments require him to adopt something like Kim's view of events and properties. Whether or not one can show that Davidson has implicitly realist commitments to properties is unclear. On the one hand, there is reason to believe Davidson's account of causation entails the claim that events have physical properties—in a realist sense—and hence, that the causal efficacy of an event is due to its real physical properties as Kim claims. On the other hand Davidson's account of mental descriptions is decidedly anti-

---

[19] Jaegwon Kim, (1976) "Events as Property Exemplifications," in Action Theory: Proceedings of the Winnipeg Conference on Human Action, ed. M. Brand and D. Walton (Dordrecht: Reidel, 1976) p. 159-77.

[20] Davidson, "Thinking Causes," p. 13.

realist, in which case he seems justified in his refusal to talk about mental properties in any realist sense.

The most compelling reason for thinking Davidson must be a realist about certain properties stems from his conception of a complete physics and from the role played by causal laws in the explanation of human behaviour. In "Actions, Reasons, and Causes"[21] Davidson popularized the view—then out of favour—that reasons explain actions because they cause them. There he defends the causal analysis of the "because" of reason-giving from the alternative analysis offered by Wittgenstein, which claims that actions are explained by fitting them into familiar patterns:

> Talk of patterns and contexts does not answer the question about how reasons explain actions, since the relevant pattern or context contains both reason and action. 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.[22]

For Davidson the explanatory force of the singular claim is ultimately derived from the underlying causal law. This works in the following way: Explanations of behaviour often invoke folk-psychological generalizations. When we explain why Peter groaned disapprovingly when he heard the punch-line by saying "The joke was awful," we understand that people tend to groan when they hear the punch-lines of awful jokes. This is an example of what Davidson calls a "heteronomic" generalization. The fact that this generalization is often true gives us "reason to believe there is a precise law at work, but one that can be stated only by shifting to a different vocabulary."[23] Again, the reason we require such a shift in our vocabulary is that our psychological concepts are unsuited to the formulation of strict laws. The idea that there is nevertheless some law at work behind the generalization is precisely what gives the explanation its explanatory force.

---

[21] Donald Davidson, "Actions, Reasons, and Causes" (1963) in Davidson.
[22] Ibid., p. 10.
[23] Davidson, "Mental Events," 219.

Davidson acknowledges the likely possibility that we might never be able to formulate the strict laws that underlie our folk-psychological explanations.[24] The fact that Davidson believes in the existence of laws we may never formulate and that the explanatory force of rationalizations is derived from such laws entails a realist attitude toward them. If these laws were not recognition-transcendent, then it would follow that in the absence of our ability to formulate strict laws causal explanations would not explain anything. This might be seen as suggesting that events stand in causal connection in virtue of their physical properties after all. Davidson thinks we should speak of predicates rather than properties, and that what properties something has is a matter of how it is described. He also thinks that laws connect events under a description; that is, laws connect events by connecting predicates. But if we admit there are laws we cannot formulate, how are the related events to be connected? It can't be in virtue of predicates, for by hypothesis we have no such predicates to connect. The most plausible answer is that the events must be connected in virtue of recognition-transcendent facts about those events, and properties (in the realist sense) do the job admirably. If this is right, then this shows there is a fact of the matter about what physical properties events have because there is a fact about the properties that figure in these causal laws. This in turn implies that there are physical properties responsible for the causal efficacy of events.

These conclusions might appear to support the epiphenomenalist objection, and many of Davidson's critics have argued along similar lines. The conclusions appear to confirm the objection because they show that Davidson cannot retain his anti-realism toward properties, which seems to grant the epiphenomenalist what he needs for his argument: The idea that events are analyzable into properties and that they stand in causal connection in virtue of certain of those properties as opposed to others (namely, the ones

---

[24] See Davidson, "Mental Events," 219 for some indication of this.

described by an ideal physics). Since only physical properties have causal efficacy, mental properties are epiphenomenal.

I think it is premature to draw such a conclusion, however, because it is unclear precisely how the "mental properties" mentioned by the epiphenomenalist fit in the discussion. We have seen there are reasons for thinking Davidson might be a realist about the physical properties of events but it is less clear that Davidson has similar commitments to mental properties. Although Davidson frequently speaks of mental properties as though they were concrete features of the world, it would seem he is actually an anti-realist about them. James Klagge has made this point recently in a related discussion about supervenience:

> The features of the mental that tend to make anomalism plausible derive from constraints upon our interpretation of other people. We ascribe beliefs and desires to people, in part, as a way of understanding, predicting, and appraising their behaviour. Thus, the mental becomes more a way of seeing people than it is something in people that can be seen.[25]

I think this is a plausible and correct way of reading Davidson, for it fits well with Davidson's general treatment of the mental. Since an individual's mental states depend on an interpretation of that individual's behaviour (by himself or others), and the states we ascribe at any one time may later be over-ruled if the interpretation requires alteration in light of new behaviour (without limit), there is no fact of the matter about someone's mental states.[26] This dependence of mental state ascription on interpretation is not unlike a

---

[25] James Klagge, "Davidson's Troubles with Supervenience," Synthese (1990) p. 342.

[26] Some might think this creates a problem for premise (1) of anomalous monism (that mental events are causally related to physical events). If there is no fact of the matter about what beliefs and desires someone has, then how can such states cause actions? The answer to this will become apparent shortly, but to anticipate we can say that characterizing what someone believes on one occasion is just the act of describing an event (probably a neurological event) in a certain way. To deny there is a fact of the matter here is just to say there is no one true mental description of that event. But since, for Davidson, the way an event is described has no impact on its causal powers, this has no consequence for premise (1).

form of anti-realism since on this view mental states become recognition- (or interpretation) dependent.

Assuming Davidson is an anti-realist about mental properties in this way, does it make sense to say, as his critics do, that on his view the mental properties of an event contribute nothing to the causal efficacy of that event? On one hand we can say it does, since a nonexistent object can have no physical effects in the world, but this is a trivial claim. Besides, if this is all we can say, Davidson's initial response to his critics holds because mental properties are then just ways of describing a particular physical event and that event will cause no matter how it is described. On the other hand it doesn't make sense to say mental properties lack causal powers because in Davidson's view there are no such properties to be epiphenomenal. This seems a more plausible thing to say because the epiphenomenalist objection is interesting only if there are actual properties in the world—in a fully realist sense—that have no causal currency. It would appear, then, that the epiphenomenalist has a case only if he can show that Davidson has realist commitments to mental properties, but it seems unlikely one could do that.

The best way to support these suggestions is with some examples, in which case it would be useful to define an epiphenomenal property. I adopt the formulation offered by Peter Hess, one of the first to raise the epiphenomenalist objection against anomalous monism. Hess's formulation is suitable because it fits the conception of epiphenomenalism implicit in what I have called "the standard objection."

> A property P is epiphenomenal with respect to the relationship between an event C and its effect E iff
> (i) P is a property of C;
> (ii) It is not the case that C would not have caused E had it not had property P.[27]

---

[27] Peter Hess, "Actions, Reasons, and Humean Causes," Analysis (1981) p. 80.

I have been suggesting there is an ambiguity about how to understand (i). P can be a property in two senses, the first corresponding to Davidson's conception, the second to Kim's. P can be a property in an anti-realist sense, meaning that it is best understood as something ascribed to C, where that ascription does not entail ontological commitments to properties because our ascription of P to C is subject to removal in accordance with whatever reinterpretations of C arise in light of new information about C and its causal exchanges with the world. Thus, C can have property P only if we recognize it as having that property; such properties are recognition-dependent. The second sense of "property" requires more serious ontological commitments than the first. In this case properties are not ascribed to events but are thought to be the ontological building-blocks of events. This means they are recognition-transcendent: C can have property P independently of us recognizing that it does.

Consider the first case where P is a property in the anti-realist sense, and let us say the events in question are identified as follows:

C=the impact of the brick against the glass.

E=the shattering of the glass .

Let us also say that John finds events like C amusing, so that we can describe C as "the event that amused John." The event C, then, has the property of amusing John:

P=the property of amusing John.

Now, is P epiphenomenal in the above sense? Of course it is. Whether or not John is amused by events like C makes no difference to the events in question. The fact that we can describe C as amusing, or in any of a number of ways has nothing to do with its causal powers, nor would we expect it to. For if John is in a foul mood and will not find anything amusing today, there is little reason to expect that the window wouldn't shatter just the same when struck with the brick. The reason for this is that the fact the impact of the brick against the glass is or is not amusing to John is not something intrinsic about the event, but

is something ascribed to the event on the basis of John's relation to it. The significance of talking about ascribed properties here is their analogy with mental properties. The property of amusing John is not an ontological constituent of the event described as "the shattering of the glass." Mental properties—on the anti-realist view—are not ontological constituents of events either. Since mental properties have the same ontological status as ascribed properties it follows that the fact that we can describe events using mental predicates is just as uninteresting for causation as the fact that we can describe event C as amusing. So the analogy shows us that the fact that an event can be picked out by a mental description says nothing about the intrinsic nature of the event and has no bearing on its causal powers. Since so-called mental properties are not constitutive of events, they are epiphenomenal in this uninteresting sense. Once again, the reason this is an uninteresting sense of "epiphenomenal" is that Davidson can say here that P is just a way of describing C (e.g. "the event that amused John"), and C will cause E no matter how we describe it. To make an interesting case against Davidson, then, his critics need to show that Davidson thinks of mental properties in the second sense.

The second sense of "property" requires that we think of properties as Kim does, as ontological parts of events and not simply as predicates, or as something that can equally be ascribed or fail to be ascribed to an event without ontological commitments. In the second sense of "property," then, P must be part of the identity of C. Let us use the example of my desiring a cold drink (together with other mental states such as believing there is a cold drink in the refrigerator, etc.) causing me to go to the refrigerator, and define our events and properties as follows:

C=my desiring a cold drink

E=my walking to the refrigerator

P=my desire for a cold drink

In this case P is a very different sort of property than in our previous example, for now P is an ontological constituent of event C; P is therefore part of C's identity. Is P epiphenomenal in the specified sense? It cannot be, and here is why: In order for P to be epiphenomenal it must be possible for event C to cause E even though it is lacking property P. But we are here assuming that P is part of C's identity. If this is the case, then had C lacked property P, C would no longer exist and so couldn't cause anything.[28] It would seem that even if one could provide compelling reasons to believe that Davidson has a realist commitment to mental properties (which I doubt), on reasonable assumptions about event-identity mental properties still cannot satisfy the conditions for being epiphenomenal.

I have tried to establish a number of things in this section. The most important was to offer an explanation for the source of debate between Davidson and his critics on whether anomalous monism leads to epiphenomenalism. I concluded that the origins of the disagreement really lie in the way Davidson and his critics characterize properties and their relations with events. The standard objection to anomalous monism presupposes a realist attitude toward properties. Davidson, however, is an anti-realist about properties in the sense that there is nothing "in" events that makes it true that they can be described in certain ways. This fundamental difference in positions has led the opposing sides of the debate to continually talk past each other, and since no one appears to have recognized this difference, the debate has reached a kind of stalemate. However, this is not to deny that Davidson's critics have a point. Indeed, my second aim was to show that there might be reason to believe Davidson actually does, or ought to, have a realist commitment to physical properties. This commitment leads to the further claim that events stand in causal connection in virtue of their physical properties, as Davidson's critics maintain. My third point was that even if we grant this claim to Davidson's critics this does not amount to a

---

[28] Cf. Peter Smith, "Hess on Reasons and Causes," Analysis (1981) pp. 208-209.

serious objection to anomalous monism. The problem is that for the epiphenomenalist objection to have any real force Davidson must also be a realist about mental properties, and in light of his holism about the mental it is unlikely one could find compelling reasons to believe this about Davidson. My final aim was to show that even if we could somehow prove that Davidson is a realist about mental properties, the epiphenomenalist argument still won't work in light of basic assumptions about event-identity. If there is an over-all conclusion to draw from these observations it is this: Those who attempt to show there is something objectionable about anomalous monism by insisting on an account of mental causation in terms of the causal role of mental properties are bound to be disappointed.

2. Reformulating the Epiphenomenalist Objection

To say that the standard objection to anomalous monism is flawed is not necessarily to deny that it has identified a real problem. On the contrary, I think the worry about epiphenomenalism is a genuine one but that we need to be more careful about how we formulate this idea. The problem with the standard form of the objection, as we saw, is that it requires assumptions about events and properties Davidson does not endorse. I believe there is another way of formulating the objection that does not make this mistake but instead casts the problem in terms Davidson himself would accept. Through a brief discussion of Dennett and Huxley I offer an alternative understanding of what it means for something to be epiphenomenal and then show how, with a few additional premises, this conception of epiphenomenalism is entailed by anomalous monism. I then identify the negative consequences this seems to have for Davidson's physicalism.

As should be evident from the previous section, the key to developing a successful epiphenomenalist challenge to anomalous monism is to avoid formulating the objection at the level of properties. Because the expression "epiphenomenal" is most often defined in

terms of properties that are causally impotent, we require an alternative understanding of this concept which nevertheless captures the idea that conscious mental states are in some sense irrelevant to the causal explanations of behaviour. Dennett has also expressed some dissatisfaction with the standard philosophical use of this term, so it will be instructive to begin this discussion with a consideration of his views.

Dennett draws our attention to the usual philosophical meaning of "epiphenomenal" which he defines as follows: "'x is epiphenomenal' means 'x is an effect but itself has no effects in the physical world whatever.'"[29] My concern about definitions of this sort is that when applied to actual cases what turns out to be epiphenomenal in this sense is a property, and this is unhelpful if we want to say that anomalous monism entails epiphenomenalism. Dennett's worry about this definition is somewhat different from mine. He complains that if this is how we understand epiphenomenalism, then either we have no reason to believe that anything fits this description, or if we do we are forced into solipsism. His reasoning is as follows: If epiphenomenal properties have no physical effects in the world at all there could be no empirical evidence for their existence, for by hypothesis there would be no way to detect them, in which case there is little reason to endorse the existence of such properties. Alternatively, if the only evidence we have for the existence of such items is private, in the sense of existing solely in our own states of consciousness (as with alleged epiphenomenal qualia), then we are forced into a kind of solipsism because if we have beliefs about our qualia those beliefs in turn can never have effects in the physical world, for otherwise qualia would no longer be epiphenomenal.[30] Dennett concludes that if this is what is meant by epiphenomenal properties we are better off believing there are no such things.

---

[29] Daniel Dennett, Consciousness Explained (Boston: Little, Brown and Company, 1991) p. 402.

[30] Ibid., pp. 402-403.

I agree with Dennett that the usual characterization of "epiphenomenal" leads to these embarrassing results and that if we are to speak meaningfully about epiphenomenalism we had better come to understand the term differently. In an attempt to construct a more plausible articulation of this concept Dennett looks back to Huxley's original treatment of the idea. (Unfortunately Dennett continues to speak of properties being epiphenomenal, but we will leave that aside for now.) According to Dennett, Huxley understood an epiphenomenal property to be "a nonfunctional property or by-product"; he says, "Huxley used the term in his discussion of the evolution of consciousness and his claim that epiphenomenal properties (like the 'whistle of a steam engine') could not be explained by natural selection."[31] Dennett claims it is perfectly consistent with Huxley's views to say that epiphenomena (like the mentioned whistle) have all sorts of physical effects in the world:

> In the same spirit, the hum of the computer is epiphenomenal, as is your shadow when you make yourself a cup of tea. Epiphenomena are mere by-products, but as such they are products with lots of effects in the world: . . . your shadow has its effects on photographic film, not to mention the slight cooling of the surfaces it spreads itself over.[32]

Dennett concludes that if we understand epiphenomenalism as Huxley did, then it is an unproblematic notion that poses no difficulties for physicalism.

Dennett might be correct about the implications of such an understanding of epiphenomenalism for physicalism, but there is reason to doubt whether or not Dennett has adequately represented Huxley's views. There is evidence to suggest that Huxley made a stronger claim than Dennett admits and this introduces an alternative characterization of epiphenomenalism that does pose a problem for physicalism, at least as it finds expression in anomalous monism.

---

[31] Ibid., p. 402.
[32] Ibid.

In his intriguing article "On the Hypothesis that Animals are Automata, and its History," Huxley conveys the results of a number of experiments that show animals can function almost normally after large sections of their brains have been removed. Indeed, Huxley's own experiments with frogs show that when prompted they are capable of performing quite complex tasks after the connections with the brain that are thought to produce consciousness have been severed:

> The frog walks, hops, swims, and goes through his gymnastic performances quite as well without consciousness, and consequently without volition, as with it; and, if a frog, in his natural state, possesses anything corresponding with what we call volition, there is no reason to think that it is anything but a concomitant of the molecular changes in the brain which form part of the series involved in the production of motion.
>
> The consciousness of brutes would appear to be related to the mechanism of their body simply as a collateral product of its working, and to be completely without any power of modifying that working as the steam-whistle which accompanies the work of a locomotive engine is without influence upon its machinery.[33]

Huxley, though, did not stop with frogs. Building on the support of a remarkable case study from Dr. Mesnet[34] involving a patient who suffered an injury to the brain, Huxley claims that the same principles hold for human beings. For as Mesnet reports, the patient, like Huxley's frog, was at times capable of quite complex, seemingly purposive behaviour without any conscious awareness at all. The conclusions Huxley draws from these observations are not nearly as tame as Dennett would have us believe. About animals Huxley says, "Their volitions do not enter into the chain of causation of their actions at all."[35] And later, when he draws the analogy between animals and humans he writes: "It seems to me that in men, as with brutes, there is no proof that any state of consciousness is

---

[33] T.H. Huxley, "On the Hypothesis that Animals are Automata, and its History" in Huxley, Methods and Results: Essays (London: Macmillan co., 1901) p. 240.

[34] Dr. E. Mesnet, Médecin de l'Hôpital Saint-Antoine, "De l'Automatisme de la Mémoire et du Souvenir, dans le Somnambulisme Pathologique," L'Union Médicale, Juillet 21 et 23, 1874.

[35] Huxley, p. 241.

the cause of change in motion of the matter of the organism."[36] This last claim is too strong to fit with Dennett's interpretation. For if states of consciousness cause no changes in the matter of the organism at all, it is hard to imagine that they could nevertheless have physical effects in the world as Dennett claims. So it would seem that it is not consistent with Huxley's views after all to say that, in the case of consciousness, epiphenomena have physical effects in the world. Dennett has not appreciated the difference between the epiphenomenalism of a steam-whistle (or shadow) and of consciousness. For while it is surely the case that the steam-whistle has a host of physical effects in the world, on Huxley's view consciousness does not.

If we understand Huxley to make this stronger claim, then our look back to the source of the concept of epiphenomenalism might not appear as helpful as Dennett suggests, for it seems as though we are back with our original understanding of the term and as we already saw such a picture of epiphenomena is extremely problematic. Our look back at Huxley has not been in vain though, for I believe there is another way of interpreting Huxley's remarks that does justice to the seriousness of his own words, yet does not lapse into the nonsense Dennett warns us against.

One way of understanding Huxley's remark that "Their volitions do not enter into the chain of causation of their actions at all" is that volitions have no role to play in the causal explanations of behaviour. In other words, conscious states do not causally explain the bodily movements of animals and humans.[37] If we say that volitions or other states of consciousness cause actions but do not causally explain bodily movement, we avoid having to deny that such states have physical effects in the world yet we retain the central meaning

---

[36] Ibid., p. 244.

[37] This might not be what Huxley really meant, but my aim here is not to determine what Huxley believed or didn't believe; I'm simply trying to offer an understanding of epiphenomena that does not collapse into the ridiculous picture rightly criticized by Dennett.

of "epiphenomena" by denying that these states have a role to play in the causal explanations of behaviour. This seems to me a characterization of epiphenomenalism worth taking seriously. What makes this idea possible is the fact that the claim that conscious mental states cause actions is partially independent of the claim that such states causally explain actions. Although it is surely the case that if reasons causally explain actions, then reasons must also cause actions, the converse does not necessarily hold.[38]

This observation allows us to distinguish two partially independent claims constituting epiphenomenalism, the first of which is stronger than the second: (a) mental states or properties have no causal efficacy, (b) identifying an agent's mental states does not provide a causal explanation of the agent's action. While (a) entails (b), the weaker claim does not entail the stronger claim. If an agent's reason does not cause his action, then identifying his reason cannot causally explain his action; however, if the identification of an agent's reason does not causally explain his action it doesn't necessarily follow that the reason didn't cause the action.

I propose to show that Davidson's anomalous monism leads to the second, weaker claim. I will achieve this in the following way. The weaker version of epiphenomenalism claims that reasons do not causally explain actions, in which case explanations citing an agent's reasons must represent a distinct category of explanation from causal explanations. Hence we require a distinction between two sorts of explanation: causal explanations and reason explanations. If I can show that Davidson's own views entail this same distinction, then I will have succeeded in showing that anomalous monism leads to the form of epiphenomenalism suggested here. Once this is achieved, I will identify the difficulties this distinction appears to create for Davidson's physicalism.

---

[38] See Simon Evnine, Donald Davidson (Stanford: Stanford University Press, 1991) p. 49.

Before I show how the distinction between these two types of explanation can be drawn from Davidson's theory it will be useful to say a few words about explanation in general. Since my task in this dissertation is not to develop an account of explanation, my remarks here will be somewhat open-ended and intuitive. Also, since my larger aim is to undermine the importance of the consequences this type of epiphenomenalism has for anomalous monism, little actually hangs on the strength of the case made for this distinction, so a definitive proof in its support is not necessary.

A promising beginning for a general account of explanation can be found in the following suggestion made by William Seager: "explanations are accounts of phenomena that aim at truth and which seek to make the phenomena intelligible to their target audience."[39] According to Seager, then, there are two conditions to be met for something to constitute an explanation: truth and intelligibility. In my view this amounts to the claim that there is an ontological component to explanation (explanations must be true to the facts in some sense) and a psychological component (we must be able to understand an explanation). I want to suggest that what is crucial to the distinction between reason explanations and causal explanations is the way the psychological condition is satisfied.

The first condition (truth) implies that a false account of some phenomenon P, will not count as an explanation of P although it might make P intelligible. Using Seager's own example, if we assume P is some "paranormal" phenomenon, ghost theory might render P intelligible to the extent we could be said to understand such a theory (indeed, I have heard people speak as though such a theory makes perfect sense to them), but since there are no such things as ghosts such a theory would offer no genuine explanation of P.[40] On the other side, a true but unintelligible account of P will not count as an explanation either. An

---

[39] William Seager, Metaphysics of Consciousness (London: Routledge, 1991) p. 18.

[40] Ibid., p. 27.

explanation that no one can understand can hardly be regarded as making the phenomenon it purports to explain intelligible, and so I see no reason to treat such an account as a genuine explanation.[41]

While this is a good start to the clarification of the idea of explanation, I think more needs to be said about the psychological component of intelligibility, so I will part company with Seager's approach at this point and offer some suggestions of my own about how to understand this idea. One way to be more specific about the condition of intelligibility is to consider cases where explanation fails. I have already mentioned the possibility that an explanation might fail because of sheer complexity. A more informative and extreme possibility can be found in those cases where the phenomenon is brute. According to contemporary science there is no explanation for why a uranium atom fissions at one time as opposed to some other time; there is simply no deeper story to tell, and hence, there is no explanation for this phenomenon. This last sort of failure of explanation is instructive because it reveals a previously unnoticed feature of explanations. If we are to have an explanation at all we must be able to refine it and tell an increasingly detailed story about the origins of the phenomenon. Explanation, then, is best thought of on a continuum: At one end we have perfectly complete explanations, explanations that can be refined in tremendous detail until we reach the point where there is nothing left to explain. On the other end of the spectrum we have brute facts for which there is no explanation, for there is nothing to refine, no deeper story to tell, not even in principle. Somewhere between we have cases where an explanation can be refined only up to a point. In this case the explanation stops not because there is nothing more to explain, but because it becomes too complicated to follow.

---

[41] For a deeper discussion of this point see Seager, pp. 17-34.

I leave it an open question as to how far an explanation must be refined until we are satisfied that an explanation has been given. Obviously this will vary depending on the level of sophistication of the target audience. However, our ordinary demands do not appear to be very high. It seems unfair to say that meteorologists have no explanation at all for the weather because they can't keep track of all of the conditions that have a bearing on the weather. Nevertheless, some degree of refinement appears necessary. For it also seems clear that if one cannot improve upon the old saying "Red sky at morning, sailor take warning," then one has explained nothing. So the question about how far an account must be refined in order to explain something remains an open one. For the purposes of my discussion it is enough to demand the minimum and say that an explanation must be refinable at least to some extent.

To sum up then, we have an explanation only if the account of some phenomenon P is true to the facts and is intelligible. An explanation is true only if it is adequate to the facts, and an account is intelligible only if it can be refined in a way that preserves understanding.

Let me now sketch out Davidson's account of causal explanation, beginning with some remarks on his view of causation. According to Davidson causation is an extensional relation, meaning that it holds between events no matter how they are described, and it is a nomological relation in the sense that underlying each singular causal claim there is a strict causal law, where the word "strict" is meant to express the idea that such laws are exceptionless. For Davidson, to explain an event is to identify its cause and, as we saw in the previous section, the explanatory force of a singular causal claim such as "a caused b" is ultimately derived from the underlying strict law. However, we need to be careful not to overstate the role played by the strict law in our explanation. Davidson urges us to resist the idea that an explanation has not been given until a strict causal law has been specified. There are two reasons for this. First, Davidson thinks it is sufficient for the purposes of explanation if our singular claim is supported by causal generalizations that are confirmed

by their instances, provided the truth of those generalizations suggests the presence of a strict law working in the background. Second, of all the singular causal claims we make in our daily traffic with events, very few of them can be sharpened into strict laws without a significant change in the vocabulary used to describe the events in question. To demand that we be able to state the covering-law relevant to any particular causal explanation is, in Davidson's view, to demand too much. So far this fits with my general account of explanation. To demand that we be able to identify the underlying strict law is akin to demanding that meteorologists be able to specify all the causal factors that have a bearing on the weather. As I said, this is too strict a demand on explanation.

Despite the fact that the explanatory force of singular causal claims inevitably lies in the underlying strict laws, in Davidson's view such laws play a decidedly background role in explanations. Given this Davidson places a great deal of emphasis on what he calls "causal lore." Our causal lore consists of the sort of causal generalizations we ordinarily appeal to when we make causal claims. For instance, we might speak of bricks typically causing windows to shatter if thrown at them in the appropriate way. These generalizations play an important role in explaining why any given occurrences of events like the first are followed by events like the second. Davidson describes such generalizations as "heteronomic" and contrasts them with "homonomic" generalizations:

> On the one hand there are generalizations whose positive instances give us reason to believe the generalization itself could be improved by adding further provisos and conditions stated in the same general vocabulary as the original generalization. Such a generalization points to the form and vocabulary of the finished law: we may say that it is a homonomic generalization. On the other hand there are generalizations which when instantiated may give us reason to believe there is a precise law at work, but one that can be stated only by shifting to a different vocabulary. We may call such generalizations heteronomic.[42]

---

[42] Donald Davidson, "Mental Events," in Davidson Essays on Actions and Events (Oxford: Clarendon Press, 1980) p. 219.

Our generalization about bricks and windows is heteronomic because in order to characterize the relevant strict law we require a change in the vocabulary used to describe the related events.

The explanations we provide for human behaviour in terms of reasons and other mental states are, according to Davidson, no different in kind from our explanations of ordinary physical events. In his view reasons causally explain behaviour and involve the same kind of heteronomic generalizations described above. For just as there is no law connecting events described as "impacts of bricks" and "the shattering of windows," there are no laws connecting events described in terms of reasons and actions (e.g., "wanting to go out for dinner" and "going to the restaurant"). The relevant covering law in the case of a reason causing an action will be formulated in other (possibly neurophysiological) terms. And when we appeal to a reason to explain an action, as with physical events we invoke a series of generalizations that are confirmed by their instances—in this case the folk-psychological generalizations connecting typical mental events to typical actions.[43] These folk-psychological generalizations suggest the presence of an underlying law and, as with our example of the brick and the window, it is from this law that our explanation ultimately derives its explanatory force. Since our explanations of ordinary physical events are causal and share the same pattern as our folk-psychological explanations of behaviour, it follows that our explanations of behaviour in terms of intentional states are also causal explanations.

Davidson has built a powerful case for the causal analysis of the "because" of reason-giving which has become the dominant view in the philosophy of action. But there is an important disanalogy between our two examples of causal explanation. In the first

---

[43] Which, it bears observing, are as folk-psychological as the talk about bricks is "folk-physical."

case (where the impact of the brick against the glass caused the shattering of the glass) we face no obstacle to a more detailed explanation of the events in physical terms. For we can refine our explanation by talking about the brick exerting a certain force and the window having a certain fragile structure and these observations, which can easily be made more precise by someone well-versed in physics, point us to a fairly complex account of the mechanisms involved in the causal relation. To put this in more Davidsonian terms, there is no principled barrier preventing us from moving from our crude heteronomic generalization to a more detailed homonomic generalization. Note, however, that in saying this I am not insisting that we need to have access to the relevant strict laws. As I said, this is to demand too much. Rather, the important thing is that the explanation can be refined or improved upon, in this case by giving us more details about the causal relation between the connected events. Since the process of refinement which I take to be essential to the concept of explanation is satisfied in this case by providing more details about the causal relation between cause and effect, this is what makes the explanation a causal explanation. I want to suggest that Davidson's views about the mental lead us to the conclusion that explanations dealing in psychological states cannot be refined in this way but are instead refinable in an alternative way, and hence, that psychological explanation represents a distinct category of explanation.

To insure that I do not beg the question against Davidson it is important to note that he himself is committed to the refinement of explanations which I take to be constitutive of the condition of intelligibility. For given the intensional character of explanation, as opposed to causation, Davidson acknowledges that we often fail to explain anything at all when the events related as cause and effect are improperly described. His example of the hurricane and the catastrophe described respectively as "the event reported on page 5 of Tuesday's Times" and "the event reported on page 13 of Wednesday's Tribune" is a case in

point.[44] To suggest that the one event explains the other under those descriptions is patently false. I want to suggest that the reason for this is that these descriptions are unsuitable as the starting point for refining the explanation, and given this intelligibility fails.

One might object that the reason the hurricane doesn't explain the catastrophe when these events are characterized in terms of newspaper headlines has nothing to do with our inability to provide a more detailed account of the relation between the mentioned events, but is instead because these descriptions of the events don't fit into a rough and ready generalization, and it is true generalizations that are important for explanations. The simple fact, however, that descriptions of events can be fit into true generalizations does not necessarily account for the explanatory force of singular instances of some generalization. It could be possible that the Times always reports events like hurricanes on a certain page of the newspaper, and that the Tribune behaves just as regularly with respect to the disasters it reports. If such were the case, then there would be a true (but perhaps accidental) generalization connecting the events under these descriptions (e.g., "events reported on page 5 of Tuesday's Times always cause events reported on page 13 of Wednesday's Tribune"), but, I think, we would be reluctant to say even then that events like the first explain events like the second when so described. Hence, fitting descriptions of events into true generalizations is insufficient to account for the explanatory force of causal explanations.

Davidson says in connection with this example that while we should not expect to find a law connecting the mentioned newspaper headlines, "[i]t is only slightly less ridiculous to look for a law relating hurricanes and catastrophes."[45] But as Louise Antony

---

[44] Donald Davidson, "Actions, Reasons, and Causes," in Davidson, p. 17.
[45] Ibid.

points out, we find explanations of disasters in terms of hurricanes perfectly acceptable.[46] Since we find explanations of disasters in terms of hurricanes to be perfectly good explanations, what distinguishes the explanation under this description from the one in terms of the newspaper headlines? Louise Antony makes a useful suggestion. She says:

> One can fruitfully inquire what sort of thing a hurricane is, with the intention of finding out how hurricanes cause disasters. If we can understand, even roughly, how things describable in the language of physics can go to make up a hurricane, we can see how the regularities describable in the language of physics can converge to produce the regularities—apparent on the macro level—that we describe in terms of hurricanes and disasters. It's the fact that we can know—even vaguely—what sort of thing a hurricane is that makes relevant the kind of strict, microphysical account of particular causal interactions that Davidson insists upon.[47]

I think that Antony is correct about this. It is simple enough to see how physical forces and the like make up something like a hurricane and how such things affect the environments in which they occur. The same cannot be said of the newspaper headlines. Thus, we can move from our description of events in terms of "hurricanes" and "catastrophes" to a more precise physical language with relative ease, and it is the possibility of this shift in precision that provides the explanation that is missing in the case of the newspaper headlines.[48] Thus, the condition of intelligibility construed as the refinement of an account of some phenomenon is crucial to the account being an explanation, even on Davidson's terms.

The disanalogy I see between straight-forward causal explanations of ordinary physical events and folk-psychological explanations is that in the case of explanations making reference to mental events we have no way of providing a more detailed account of

---

[46] Louise Antony, "Anomalous Monism and the Problem of Explanatory Force," Philosophical Review Vol. XCVIII (April, 1989) p. 169.

[47] Ibid., p. 170.

[48] Of course the difficulty is not necessarily one of principle in the case of the generalized newspaper headlines. One could presumably research the reports and see the connections between them in more salient terms. But this makes precisely the needed point. We must, if we are to have a genuine causal explanation, be able to refine our explanation at least to the point where we can appreciate how the one event brought about the other, even if such appreciation involves a fairly crude understanding.

the causal relation connecting reason and action. That is, unlike physical forces and hurricanes we have no way of seeing how mental phenomena (such as having reasons) cause actions, which means that the nature of the particular connection between a reason and an action is completely inaccessible to us. This inaccessibility is a unique feature of our folk-psychological explanations because for Davidson it is one that holds in principle—and not just because of sheer complexity. Because of Davidson's belief in the anomalism of the mental we cannot move from our folk-psychological generalization to a physically more precise vocabulary that informs us about the causal connection between the related events as we could with our previous examples. Davidson expresses this idea quite starkly: "Even if someone knew the entire physical history of the world, and every mental event were identical with a physical, it would not follow that he could predict or explain a single mental event."[49] Given this we cannot take even the first step in refining our explanation, for we have no way to shift from our psychological vocabulary to the more precise physical vocabulary that would give us the deeper details about the nature of the connection between the reason and the action.

Louise Antony arrives at the same conclusion about Davidson's treatment of psychological explanation: "Davidson cannot explain the explanatory force of the rational 'because' by appeal to underlying causal connections because there is no objective attachment between the interpretive psychological story we decide to tell and the physiological goings-on in a person's body."[50] Although Antony's reasons for denying there can be a connection between mental and physical descriptions are slightly different from those I have emphasized, both stem from the anomalism of the mental and the barrier this creates to our understanding of the causal connections between reasons and actions.

---

[49] Davidson, "Mental Events," p. 224.
[50] Antony, "Anomalous Monism and the Problem of Explanatory Force," p. 184.

The conclusion reached is the same in either case. Without a means of connecting reason explanations to an account of the causal mechanism between reasons and actions, there is nothing to account for their explanatory force if such explanations are taken to be a species of causal explanation.

It seems quite obvious that reasons nevertheless do explain actions, but if they do not causally explain actions then how do they explain them? As Davidson says: "If reason and action illustrate a different pattern of explanation, that pattern must be identified."[51] The alternative "pattern" is perfectly obvious and fits nicely with Davidson's own treatment of the mental. As has been suggested many times, reasons explain actions by showing them to be rational. Our practice of doing this to explain the behaviour of others typically takes the form of ascribing certain beliefs and desires to agents in accord with other things we know about them. Through such ascriptions we show how the agent's behaviour is rational in the light of the agent's other beliefs, desires, and the like. By situating actions within the normative constraints of rationality in this way we ordinarily feel perfectly satisfied that the behaviour in question was explained.[52] Thus, in this case the condition of intelligibility is satisfied in a different way than it was for the causal explanations discussed earlier. The difference is in how the explanation is refined. With a causal account the explanation is refined by gaining an appreciation—no matter how basic—for the causal mechanism at work between the events in question; by contrast, in this case the explanation is refined by identifying additional mental states which situate the behaviour in a normative context. Because the nature of the refinement is radically different we have a different species of explanation.

---

[51] Davidson, "Actions, Reasons, and Causes," p. 10.

[52] Jaegwon Kim suggests a similar account of how reasons explain actions which, he claims, is consistent with "psychological anomalism" in his paper "Self-Understanding and Rationalizing Explanations," Philosophia Naturalis 21 (1984) pp. 309-320. I am in basic agreement with the more detailed account he offers there.

It is important to note that our use of this procedure does not require a belief in anything as contentious as "ideal rationality." For the holistic character of belief ascription suggested by Davidson has the advantage of allowing us a fairly broad (as opposed to idealized) notion of rationality which permits us to explain behaviour that might at times even appear irrational. Suppose a patient walks into a doctor's office and demands that the doctor operate on his knees even though there is nothing physically wrong with them. At first this behaviour seems inexplicable. However, given that the patient believes he is under surveillance by the CIA, and that he thinks they have implanted homing devices in his knees, and given that he also wishes to escape his imagined observers, this behaviour is explicable after all, and it is explainable by showing how it is rational against the background of the agent's other mental states. Since the concept of reason plays a central role in such explanations I call them "reason explanations."[53]

This example raises an interesting question. On my account of the explanation of behaviour, how are we to distinguish between genuinely informative reason explanations and mere rationalizations[54] in the pejorative sense (the making-up of reasons or excuses)? Is there room for such a distinction in my account? It seems there ought to be, if the account is a good one, because we ordinarily think that the distinction is a correct and useful one to make. We do as a matter of fact distinguish between cases where we think that someone has provided a genuine explanation of his or her action and those cases where the explanation is plausible and fits with the available evidence but we have been deceived about the agent's reasons. Since the central feature of my account of the explanation of

---

[53] Peter Hess and Mark Thornton have also insisted that such explanations are not causal explanations, and refer to them by similar labels. See Peter Hess, Thought and Experience (Toronto: University of Toronto Press, 1988); Mark Thornton, Folk Psychology: An Introduction (Toronto: Canadian Philosophical Monographs, 1989).

[54] Davidson uses the term "rationalization" in the non-pejorative sense I call "reason-giving." To avoid confusion in the ensuing discussion "rationalization" should not be taken in the Davidsonian sense, but rather in its more common pejorative sense.

behaviour is that the behaviour is shown to be rational against the background of the agent's other mental states, it is unclear that I can make this distinction, for both explanations render the action rational.

Given Davidson's holism about the mental it might appear as though there is no distinction to be made between genuine reason explanations and rationalizations. For it might also seem to follow from the thesis about holism that there is no fact of the matter about what reasons someone has for performing an action.[55] Because the distinction between these kinds of explanation is usually explained in terms of what reasons someone "really had" for acting, it appears this distinction makes no real sense, in which case the difficulty here is not real, but only apparent. So long as the allegedly manufactured reasons cohere with all the available evidence (such as the agent's other beliefs, desires, and actions) as well as so-called real reasons, we have no reason to prefer the one interpretation of the agent's action over the other.

It might be that what allows us to identify rationalizations as such is just that they never do cohere with the agent's other mental states and behaviour as well as genuine explanations do, in which case we can make the distinction between these two types of explanation without any difficulty. However, this might be an unwarranted assumption. I see no reason in principle why we should deny the possibility that someone who is very clever should be able to provide us with a manufactured story about why she performed some action which coheres with all the available evidence as well as some other account of why she did what she did. In this case how are we to make the distinction between her real reasons for acting and the ones she invented?

---

[55] Louise Antony suggests this toward the end of "Anomalous Monism and the Problem of Explanatory Force," p. 184.

The route to making this distinction lies in a related discussion in "Actions, Reasons, and Causes." There Davidson responds to those who deny that reasons are causes of actions. He claims that something that should worry his opponents is the fact that an agent can have a reason for performing an action yet not perform the action for that reason. Since the apparent reason and the "real" unidentified reason both render the action rational, how are we to make sense of the notion that there is a particular reason for which the agent performed the action? The analogy here with the problem about rationalizations should be perfectly obvious. Davidson's answer is that the "real" reason is the one that caused the action.[56] The same reply solves the problem about distinguishing between genuine explanations and rationalizations.

Can I make use of the same maneuver? I believe so. I have denied that reasons causally explain actions. This should not be confused with a denial that reasons cause actions. I have not denied this. As I pointed out earlier, these two claims are partially independent. The fact that the identification of reasons does not causally explain actions need not entail that reasons do not cause actions. So the view of reason giving we arrive at is this: In the case of a genuine explanation we identify the reason that caused the action and renders it rational; in the case of a rationalization we identify a reason that renders the action rational but did not cause it.[57]

This distinction between the reason that caused an action and one that merely gives an action the appearance of rationality brings in Seager's first criterion for explanation: truth. For we are committed to the claim that there is a fact about which reason caused an action, and if our reason explanation is to explain anything at all we must correctly identify that reason. Thus, reason explanations (when we get them right) seem to satisfy Seager's

---

[56] Davidson, "Actions, Reasons, and Causes," pp. 8-11.

[57] How we can recognize the difference is an interesting and difficult question which is worth pursuing, but unfortunately goes well beyond the scope of this paper.

criteria for being proper explanations. The act of rendering behaviour rational satisfies the intelligibility requirement because such explanations can be refined according to general principles of rationality, and the claim that there is a particular reason that caused an action satisfies the truth requirement. Because the intelligibility requirement can be satisfied only on a non-causal construal of the connection between a reason and an action, explanations of this type are not causal in nature. Hence, reason explanations are not a species of causal explanation.

Since, given some reasonable demands on the nature of explanation it follows from Davidson's anomalous monism that reasons do not causally explain actions, anomalous monism leads to one of the claims of epiphenomenalism. However, as I pointed out this is the weaker of two claims. The weaker claim is significantly different from the stronger claim in several respects, which is generally in Davidson's favour. This claim is not equivalent to saying that mental events or properties are causally impotent, for the claim that reasons do not causally explain actions is consistent with the claim that our intentional states have a causal role to play in the production of our behaviour, and so we retain all the benefits of Davidson's thesis that reasons are causes of actions. However, given these observations one might wonder what the significance is of showing that anomalous monism leads to the weaker claim. In particular, what aspects of Davidson's theory does it affect?

Davidson thinks that reasons causally explain actions, so in light of the above this claim is something that needs rethinking in his theory. But the significance of this appears to run deeper than merely demanding that Davidson revise his account of psychological explanation. By denying that reasons causally explain actions reason explanations emerge as a distinct type of explanation that appears to be autonomous, for such explanations do not (given the arguments mentioned) seem to be connected or connectable to causal explanations that deal in physical predicates. Thus, it appears as though this distinction

between types of explanation entails that psychological explanation is autonomous. If this is indeed the case, then these considerations undermine physicalism's claim to explanatory completeness. This is a fatal blow to Davidson since he should be regarded as accepting the belief that physicalism is explanatorily complete, despite his antireductionism. His allegiance to this idea is illustrated by the very fact that he treats reason explanations as causal explanations and locates the source of their explanatory power in underlying physical laws. By insisting that the capacity reasons have for explaining actions is derived from physical laws, Davidson incorporates physicalism's claim of providing complete explanations for all phenomena into his theory. By showing that Davidson's anomalous monism leads to the weaker epiphenomenalist claim I have cast doubt on whether anomalous monism actually supports this aspect of physicalism.

In conclusion, I have shown that the nagging suspicion felt by many philosophers that anomalous monism leads to some form of epiphenomenalism is justified. However, the form this objection takes is somewhat different than most have thought. The traditional worry was that on Davidson's account of causation mental properties have no causal role to play in the production of behaviour, nor, it would seem, do they have any physical effects in the world at all. As we saw, this way of putting things is problematic in two respects. First and foremost, this version of epiphenomenalism requires a realist commitment to mental properties Davidson does not endorse. Second, as Dennett has shown, this understanding of epiphenomena lapses into utter nonsense. Through a look back to Huxley's original thoughts on epiphenomena I differentiated a weaker understanding of the concept as the claim that while conscious mental states cause behaviour, the identification of such states in psychological explanations does not causally explain behaviour. By considering Davidson's account of causal explanation and the anomalism of the mental I showed that Davidson's views actually entail this weaker version of epiphenomenalism and that this appears to entail that reason explanations are autonomous, which in turn has

unfortunate consequences for his physicalism. As we shall see later, however, there are good reasons to doubt that the conclusion that reason explanations are autonomous actually follows from the distinction between reason explanations and causal explanations, in which case Davidson's theory is vindicated.

Chapter 3

The Problem of Qualia

Some philosophers regard qualia as representing a fundamental challenge to physicalism. For example, William Seager claims that "the problem of qualia . . . emerges as the clearest threat to even . . . [a] minimal physicalism."[1] The term "qualia" refers to what is variously described as "raw feels," "phenomenal properties," "subjective experiences," and the like; they are the way conscious experiences feel to those who have them. For instance, qualia are what it is like subjectively to feel a pain, to see something red, to taste something bitter, and so on. The difficulty qualia pose for physicalism typically takes one of two forms: either qualia generate problems for particular forms of physicalism, generally identity theories, or they appear to escape capture in physical terms entirely, in which case physicalism begins to seem incomplete in some way. The conclusion typically drawn from such difficulties is that qualia are epiphenomenal properties. My aim in this chapter is to explore and evaluate the two ways in which considerations of qualia have created problems for physicalism.

In the first section I examine the problem that qualia pose for one particular variety of physicalism known as "functionalism." Although functionalism is officially neutral with respect to the ontological status of the mind, it is usually regarded as an articulation of physicalism and might best be characterized as a more advanced version of the type-identity

---

[1] William Seager, Metaphysics of Consciousness (London: Routledge, 1991) p. 38.

theory. Recall from the previous chapter that one of the major shortcomings with the type-identity theory is that it is not flexible enough to allow for the multiple realizability of mental states. Since, for the type-identity theory, mental types such as pains are identical with types of brain states the theory does not allow for the possibility that other forms of life (Martians, silicon-based life forms, etc.) with radically different kinds of nervous systems or analogous structures might also feel pain. If it is possible that such creatures feel pain but are physiologically quite different from us, then pains cannot be identical with types of human brain states. The improvement of functionalism is that it allows for much more flexible characterizations of physical types which can be identified with mental states in a less problematic manner. This flexibility is made possible by the fact that the physical type involved is a functional type defined by its causal role in an organism. The functional state of an organism is a relational state defined in terms of its causal role with respect to inputs (stimuli), outputs (behaviour), and other "mental" states (also functionally characterized). The benefit of this approach is that since mental states are functionally defined they can be attributed to organisms with radically different physical structures. Thus, the main virtue of functionalism is that it addresses the problem of multiple realizability.

The main difficulty with functionalism is that it seems to be incomplete in the way it defines mental states. This problem is most evident in the case of the inverted spectrum argument. This objection introduces the possibility of inverted qualia. Thus, some philosophers have argued that it is possible for a functional state identified as the perception of red, for example, to be accompanied by a green quale. That is, although you and I might consistently identify the same objects as green or red, I have the sort of experience looking at red objects that you have when looking at green objects. Since, by hypothesis, there are no behavioural cues to betray the difference in the phenomenal character of the experience for the person perceiving the inverted spectrum, it is possible to misidentify the mentioned

mental state functionally, which means that the mental state cannot be identical to a functional state of the system in question.

The interesting feature about this objection is that it points to the incompleteness of physicalism. If the objection is a good one, then this particular version of physicalism is incomplete, for there are mental states that are not captured functionally. My intention is to show that the argument against functionalism based on a hypothetical qualia inversion is unconvincing because more detailed examination of this idea reveals that an undetectable qualia inversion does not represent a plausible hypothesis.

In the second section of this chapter I examine an argument which attempts to refute physicalism directly (regardless of its particular formulation) by showing that there are truths about the experiences of others that cannot be captured by physical facts (no matter how such facts are to be construed). These "facts" are the same ones that cause problems for functionalism: they are facts about the subjective character of experience. Since physicalism cannot account for such facts the physical sciences are said to "leave something out." In other words, there are more facts to capture than the physical facts, which means that physicalism is false. Although similar, this is a distinct objection from the inverted spectrum argument. The point here is not that two functionally identical systems might possess different qualia; rather, the point is that there are facts about human experience that cannot be captured by any physicalist account of the mind. The argument is Frank Jackson's "knowledge argument" developed in his paper "Epiphenomenal Qualia."[2] His bold denial of physicalism has attracted a great deal of attention and criticism.[3] My aim is to

---

[2] Frank Jackson, "Epiphenomenal Qualia," Philosophical Quarterly, 32 (1982) pp. 127-136.

[3] For a sampling of these see Terence Horgan, "Jackson on Physical Information and Qualia," Philosophical Quarterly, 34 (1984) pp. 147-152; Paul Churchland, "Reduction, Qualia, and the Direct Introspection of Brain States," Journal of Philosophy, 82 (1985) pp. 8-28; David Lewis, (1983) "Postscript to "Mad Pain and Martian Pain"," in Lewis, Philosophical Papers Vol. 1 (Oxford: Oxford University Press, 1983); Laurence

assess the relative strengths and weaknesses of Jackson's argument and the responses to it. As we shall see, the actual force of Jackson's objection to physicalism requires some clarification since he appears to draw implausibly strong ontological conclusions which his critics rightly think are unwarranted. However, I see the possibility of deriving a weaker conclusion from Jackson's argument that once again appears to point to the explanatory incompleteness of physicalism.

1. Inverted Qualia

Functionalism is the view that mental states are identical with certain organizational states of a system. Typically such states are defined in terms of the causal relations between stimuli, other mental states, and behaviour. This view has been the dominant theory of mind for the past twenty years or so, emerging alongside powerful research programs in cognitive science and artificial intelligence. Functionalism is therefore a view that philosophers of mind who criticize physicalism (in its various incarnations) must deal with at one point or another. One of the most infamous and controversial arguments against functionalism is the "inverted spectrum" argument. The argument claims it is possible for functionally equivalent systems to have different colour qualia. For instance, one individual might have the sort of experience when looking at red things that another individual has when looking at green things, even though their behaviour toward coloured objects is indistinguishable (e.g., they make all the same colour discriminations). This is not an idea that was thought up just to create difficulties for functionalism. In fact, the conjecture that such an inversion is possible goes back at least as far as Locke, who in his Essay asks us to consider the possibility that "the same Object should produce in several Men's Minds different Ideas at

---

Nemirow, "Physicalism and the Cognitive Role of Acquaintance," in Mind and Cognition: A Reader, ed. W. Lycan (Cambridge: Cambridge University Press, 1990).

the same time; e.g. the Idea, that a Violet produces in one man's Mind by his Eyes, were the same that a Marigold produced in another Man's, and vice versa."[4] Assuming it is possible for such inversions to occur they are a problem for functionalism only if the qualitative differences do not manifest themselves in behaviour, and we might take Locke's conjecture to assume that such differences could exist without ever being detected.[5] The reason such inversions pose a problem for functionalism is that it seems as though we could have a case where two individuals are functionally equivalent yet differ in their mental states. This means that not all mental states are identical to functional states, and hence, that functionalism is false. Furthermore, since these qualitative differences are not captured by the causal relations characterizing functional states it would seem that qualia are epiphenomenal, for in this case they play no causal role.

The very possibility of qualia inversion has been hotly debated.[6] For the most part those who object to the possibility of Locke's conjecture have contested the inversion on verificationist grounds, claiming either that mental differences that do not manifest themselves in behaviour are no differences at all, or that the hypothesis invokes the notion of a private language which is nonsensical. Thus, responses typically attempt to prove that such an inversion is impossible on a priori grounds. I propose a different approach to the evaluation of this idea. I think that attempts to show that spectrum inversion is impossible

---

[4] John Locke, An Essay Concerning Human Understanding, ed. Peter H. Nidditch (New York: Oxford University Press, 1975) p. 389 (Bk. II, Ch. XXXII, sec. 15).

[5] Locke's conjecture is somewhat different from the inverted spectrum hypothesis since Locke simply pointed to the possibility of qualitative differences in very limited cases, whereas the inverted spectrum hypothesis requires a systematic inversion of all colour qualia.

[6] For what is probably the most famous exchange of views on this, see: Ned Block, (1978) "Troubles with Functionalism," in Readings in Philosophy of Psychology Vol. 1, ed. Block (Cambridge: Harvard University Press, 1980) pp. 268-305; "Are Absent Qualia Impossible?" Philosophical Review LXXXIX (1980) pp. 257-274; Sydney Shoemaker, (1975) "Functionalism and Qualia," in Block, pp. 251-267; "Absent Qualia Are Impossible-A Reply To Block," Philosophical Review XC (1981) pp. 581-599; "The Inverted Spectrum," Journal of Philosophy LXXIX (1982) pp. 357-381.

aim at too strong a conclusion. For we can admit the logical possibility of undetectable qualia inversions, but it seems odd to think that the mere logical possibility of spectral inversion should defeat functionalism. After all, dualism is a logical possibility too, but we tend not to think that this undermines functionalism. Thus, the real issue when speaking about spectrum inversion should not necessarily be one of mere logical possibility, but of plausibility. It seems to me that it is enough, at least for my purposes here, to give reasons for thinking that the hypothesis of qualia inversion is false, even if not impossible. To achieve this I will show that it is reasonable to believe that any systematic inversion of phenomenal elements in experience will, contrary to the modern form of Locke's conjecture, lead to differences in behaviour.[7] This is because there are intimate connections between qualia and affective states, such that if qualia are inverted this necessitates an inversion of the connected emotional states, and such inversions will inevitably manifest themselves in behaviour. The resulting behavioural differences betray the underlying phenomenal differences and so the latter can be functionally captured. I will argue this directly in the case of colour qualia inversion and will then show that similar problems hold for the qualia associated with the other sensory modalities. The examination of connections between affect and sensation in other cases will serve not only to reinforce the conclusion drawn in the original case of vision, but will also rule out the likelihood of qualia inversion in the other sensory modalities. This exercise will show that the inverted spectrum argument is one that we should not take seriously.

---

[7] Jean Harvey also argues that any inversion of phenomenal elements would be detectable, but does so on the basis of an implausibly strong assumption about type-type correlations between neurophysiological and mental states. Since this would appear to rule out anomalous monism as a plausible theory of mind, I will argue along quite different lines. For Harvey's argument see her "Systematic Transposition of Colours," Australasian Journal of Philosophy 57 (1979) pp. 211-219.

It is, I think, widely recognized that pain is often associated with quite strong emotional reactions. In fact, the associations often appear to be so strong that one might think that they are necessarily connected. In "Sensation Deconstructed,"[8] Evan Simpson develops this idea and employs an analogy between sensations of pain and perceptions of danger to show that emotional content partly constitutes sensations of pain. He suggests that just as fear is a necessary affective component to the perception of danger, self-pity is a necessary component to the sensation of pain:

> Pain, too, has a conceptual structure. For something to seem painful is for it to be experienced pityingly, just as for something to seem dangerous is to experience it fearfully. Pity is not only a reflection upon suffering but also a conceptually constitutive factor in it. When we do not feel sorry for ourselves, I suggest, we are not in distress, though there are many instances in which this self-pity is tempered with the knowledge that the suffering must be borne without complaint.[9]

So just as the perception of danger evaporates when there is no longer any sense of fear, the discomfort of pain disappears when there is no sense of self-pity. The emotional state is therefore partly constitutive of the experience.

While I think there is room for debate about whether or not self-pity is the correct emotional state to focus on, the general idea seems to me a plausible one and is in fact supported by a great deal of medical research. For instance, the International Association for the Study of Pain defines pain in the following way: "it is unquestionably a sensation in a part or parts of the body but is also always unpleasant and therefore also an emotional experience."[10] While this is a philosophically contentious claim, not least because of the assumption that whatever is unpleasant is an emotional experience, the identified

---

[8] Evan Simpson, "Sensation Deconstructed," in Entities and Individuation: Studies in Ontology and Language in Honour of Neil Wilson, ed. Donald Stewart (New York: Edwin Mellin Press, 1989).

[9] Ibid., p. 162.

[10] International Association for the Study of Pain, "Classification of Chronic Pain: Descriptions of Chronic Pain Syndromes and Definition of Pain Terms," Pain Suppl. 3 (1986) S217.

connection between pain and emotion is very suggestive and supports Simpson's assimilation of these two elements. This connection finds additional support in further research. For example, a recent article by Ephrem Fernandez and Dennis C. Turk[11] suggests that a variety of studies show that although the sensory and affective components of pain may be conceptually separable they are not independent in the sense of being fully detachable. Thus, regardless of whether the processing of the sensory component is parallel to the processing of the affective component, or follows upon it, there are, in their view, two aspects of pain that are necessarily connected in experience: the sensory and the affective.[12]

Simpson suggests that this model can be assimilated to other sensations. In his view, not only do sensations of pain have an affective component which is partly constitutive of the sensation, but the same can be said of sensations of colour:

> We well know . . . that our attention is drawn to certain colors, such as the blue of the sky on a fine day and the pure colors as refracted by a prism. Such colors impress us as worthy of attention, as do some pale and mixed colors once we have developed a little sophistication. Judgments of interest-worthiness might seem less closely bound to perceptions of color than judgments of pitiability are to perceptions of pain, since the perception of color remains even in uninteresting cases—unlike pain judged as trivial. But this line of thought is dubious: pain which lacks its normal affective character can still be identified, in the sense that one knows how to answer the question, "Where does it hurt?" Similarly, we can answer the question "What color is it," in cases where the color is without interest or attraction for us . . . .[13]

Simpson's aim is to show that given the assimilation of the structural features of sensations to emotions, and given the communicability of emotions, we have little reason to expect that there is anything incommunicable in sensation, any ineffable content that could be

---

[11] Ephrem Fernandez and Dennis C. Turk, "Sensory and Affective Components of Pain: Separation and Synthesis," Psychological Bulletin 112, No. 2 (1992) pp. 205-217.

[12] Some researchers claim that pain and emotion are processed separately in parallel. See H. Levenhal and D. Everhart, "Emotion, Pain, and Physical Illness," in Emotions in Personality and Psychopathology, ed. C. E. Izard (New York: Plenum Press, 1979) pp. 263-279.

[13] Evan Simpson, "Sensation Deconstructed," p. 163.

inverted without our expressing the difference. This defeats Locke's conjecture that there could be an undetectable inversion of colour qualities and so can also be used to salvage functionalism from the inverted spectrum argument.

I think that Simpson's approach is very promising, but as he himself recognizes, his arguments for the assimilation of sensation to emotion are suggestive rather than deductively tight.[14] One would have to make a fuller case for the connection between sensation and emotion in order to show that there is a strong analogy between sensations of pain and other sensations, such that the others also have non-detachable affective contents. If one could supplement his suggestions with some hard evidence, as I did above for the connection initially suggested between pain and affective content, his argument against colour qualia inversion will be significantly stronger. This is my goal in what follows. I begin by surveying evidence for this connection in the case of colour sensation and then move on to other cases. In each instance the conclusion is the same: there are intimate connections between sensation and emotion for all sensory modalities, and in each case it appears as though the possibility of undetectable qualia inversion is extremely unlikely as a result of these connections. This discussion should give us sufficient reason to think that the inverted spectrum argument and its analogues for the other senses do not represent serious possibilities.

Starting with vision, evidence for the connection between sensation and affect is readily available. We have all heard the results of research in environmental psychology which claims to show that colours have quite specific emotional connections. It is also common knowledge that interior designers regularly make use of these results when creating the environments in which we frequently find ourselves. Diane Ackerman describes several of these connections between colour sensation and emotion:

---

[14] Ibid., 163.

> Children will use dark colors to express their sadness when they're painting, bright colors to express their happiness. A room painted bubble-gum pink (known in hospitals, schools and other institutions as "passive pink") will quiet them if they've gotten obstreperous. In a study done at the University of Texas, subjects watched colored lights as their hand-grip strength was measured. When they looked at red light, which excites the brain, their grip became 13.5 percent stronger. In another study, when hospital patients with tremors watched blue light, which calms the brain, their tremors lessened. Ancient cultures . . . used color therapies of many sorts, prescribing colors for various distresses of the body and soul. Colors can alarm, excite, calm, uplift. Waiting rooms in television studios and theaters have come to be called greenrooms, and are painted green because the color has a restful effect.[15]

There is therefore a well researched body of evidence pointing to the connection between affect and sensation in the case of colour vision. While I suppose the research is ambiguous about whether these emotional reactions are effects of colour experiences or are part of the experiences themselves, as Simpson suggests, I think the intimacy of the relation says a lot for Simpson's hypothesis. At any rate, we should hardly find it surprising that the psychologists conducting research in this area have not characterized this connection in the philosophical manner Simpson does. Their interests are not philosophically driven.

If we accept the connection between affect and sensation—and the empirical research mentioned above supports this for sensations of colour—then we have every reason to believe that colour qualia inversion, were it to happen, is detectable. For if colour qualia are inverted, then it follows that their affective components must also be inverted since these are not radically separable. This means that subjects with inverted colour qualia will respond differently to tests designed to capture their emotional responses. For example, we would expect such people to become increasingly excitable in so-called greenrooms or rooms painted "passive pink," and more sedate in red environments. While this is probably over-simplifying the matter, there is every reason to believe that although

---

[15] Diane Ackerman, A Natural History of the Senses (New York: Vintage Books, 1990) pp. 254-255.

someone with inverted colour qualia might have learned to say that passive pink is "relaxing" and that red is "stimulating," psychologists could design elaborate tests to see past such verbal reports and determine genuine emotional reactions to perceived colours. Thus, there is good reason to believe that colour qualia inversion will lead to behavioural differences, and hence, that such differences can be functionally captured.

The case against colour qualia inversion is therefore quite strong. Of course the argument does not establish the impossibility of Locke's conjecture, but that wasn't the goal. Still, I think the conclusion can be reinforced if we examine the other senses and show that, despite the initial plausibility of analogous inversions of phenomenal elements for those senses, there are similarly strong connections between sensation and affect which remove this air of plausibility. Thus, what follows is a kind of argument from analogy to show not only that the connection between sensation and emotion is a plausible general hypothesis, but that it also undermines the inversion hypothesis for other qualia. The larger aim here is to provide a richer account of a physicalist view of sensation, at least in contrast to Locke's. The fuller articulation of the nature of qualia will make the physicalist view of phenomenal properties more plausible.

The idea that there are affective components that are partly constitutive of all sensations is not such a new one. C. S. Sherrington suggested this connection between affect and sensation some time ago in a medical context.[16] There is therefore already some support for this view, but as we shall see there is no need to rely on medical research alone. In tracing these connections in the other senses it is worth noting at the start what features of colour qualia lend the inversion hypothesis its initial credibility. Once this is clear we can then focus on the phenomenal elements associated with the other senses that share these

---

[16] See C. S. Sherrington, "Cutaneous Sensations," in Textbook of Physiology, ed. E. A. Schafer (London: Pentland, 1900) pp. 920-1001.

features. By making these cases as analogous as possible I hope to lend further plausibility to the idea that most qualia are structurally similar, in which case it is more likely that they share the characteristic of having affective contents.

Of central importance to the inverted spectrum hypothesis is the fact that the perceived colour spectrum is a structured continuum, though with definable elements (primary colours), beginning (for us) with red and ending in violet. The elements therefore have a fixed sequence in "colour space" and hence are ordered, allowing us to conceive of the inversion in terms of flipping the colour spectrum to run from violet to red. Since it is important for the inversion hypothesis that the relations between the phenomenal elements be preserved, the fact that there are structural relations at all between the elements, including the fact that they can combine to produce a tremendous number of complexes, is also central. Finally, it is important to note that there is no tight conceptual connection between colour qualia and the nature of the surfaces that cause them, for although we understand that the colours we see are caused by physical properties of surfaces (e.g., their propensity for reflecting and absorbing certain wavelengths of light), it is not immediately obvious that a particular surface will cause colour sensations of a certain sort.

I will begin the examination of the other sensory modalities with the sense of hearing. The most likely of auditory phenomenal elements to lend themselves to inversion appear to be those of tone and pitch. This seems plausible because there is a strong analogy between such elements and colours. For we have an ordered continuum of phenomenal elements (like musical notes) that share relations with one another such as harmony and dissonance and can combine to form complexes, which can also be said of colours. There is also the fact that colours and sounds are both caused by waves, and hence there is no strong conceptual connection to be made between the phenomenal properties and their causes. Further evidence in favour of the analogy is the fact that "colour" can refer to tonal quality in music.

Someone with inverted auditory qualia, call her "Jane," would hear pitch the opposite way we do, so that high notes sound like low notes, and vice versa. Such a difference will not be immediately detectable. For instance, we could imagine Jane being a member of a choir. Because she hears C sharp as a much lower note than the rest of the choir doesn't mean that she will sing off key. For she will hear that same sound when she herself sings C sharp. Thus, although the music sounds completely different to Jane, she will not sing any differently from the rest of the group because the notes she sings must match the way the rest of the choir sounds to her. However, when we consider the way most people process music and the emotional associations we tend to make with certain patterns of sound, it becomes very difficult to imagine that Jane's inversion could pass undetected.

There are very few people who have perfect pitch, and thus it is generally thought that music processing is completely relational. In recognizing a melody, for instance, one does not so much recognize the individual notes or their sequences, but rather the melody's contour (ups and downs in pitch).[17] It has been suggested in a number of different contexts that there are natural connections between contour and emotion. Speech-prosody work by Ann Fernald supports the claim that some speech contours in infant-directed discourse are present across cultures and share the same emotional meaning (e.g., down-up for soothing).[18] Also, following work by M. M. Lewis, she suggests that the affective aspect of contour plays an important role in the development of linguistic understanding:

> As Lewis observed many years ago, when "we consider the child's response to speech we must recognize that apart from its expressive functions and conventional meaning it will have an effect upon him merely because of its musical and affective

---

[17] My thanks to Glenn Schellenberg for sharing his insights into the psychology of music.

[18] See Anne Fernald, "Intonation and Communicative Intent in Mothers' Speech," Child Development 60 (December 1989) pp. 1497-1510; "Prosody and Focus in Speech to Infants and Adults," Developmental Psychology 27 (March, 1991) pp. 209-221.

> qualities"... It is through these "musical and affective qualities," according to Lewis, that speech first becomes meaningful to the infant.[19]

In philosophy the connection between music and emotion has long been emphasized in aesthetics. While theories vary about the details, there appears to be strong agreement that certain musical contours have quite specific emotional connections. For example, Peter Kivy claims that certain patterns of sound are expressive of specific types of emotions (i.e., call to mind particular emotional states, rather than express the feelings of the composer). The down-up contour in music (as in speech-prosody) expresses elation and the minor triad is expressive of "dark" emotion. While he correctly acknowledges that convention plays a large role in such associations, he thinks that the distinction between convention and natural association collapses in the end.

> It is inviting to suppose that many musical features expressive by convention were once more than that: were once heard as resembling identifiable expressive behavior, or at least ingredients in such structures. I am inclined to believe it is the case.[20]

While I do not here wish to commit myself to any particular aesthetic theory of music or theory of language acquisition, the general claim that there are natural connections between contour and emotion seems quite plausible and can be used both to reinforce the thesis that affect is partly constitutive of sensation, and to argue against an undetectable inversion of auditory qualia. The intimacy and apparent universality of the connections between contour and emotion are strong evidence for the first point. Granting that such connections hold, an undetectable inversion of auditory elements seems most implausible. For if something as basic as the down-up contour has a natural emotional association, then

---

[19] Anne Fernald, "Intonation and Communicative Intent in Mothers' Speech," Child Development 60 (December 1989) p. 1508. See also M. M. Lewis, Infant Speech: A Study of the Beginnings of Language (London, Routledge & Kegan Paul, 1936/1951) p. 44.

[20] Peter Kivy, The Corded Shell: Reflections on Musical Expression (Princeton: Princeton University Press, 1980) p. 82.

(as we did for colour perception) we can imagine performing tests on individuals designed to determine their emotional reactions to contour and determine if there are any consistent, but unexpected differences in emotional response. If Jane's auditory qualia are inverted, then she will hear the down-up contour as the up-down contour and consequently her emotional response should be different from the norm. The consistent difference in Jane's behaviour points us to the fact that the nature of her auditory experience is different, and since the phenomenal difference leads to a behavioural difference, the phenomenal difference can be functionally captured. It appears, then, that the inversion hypothesis cannot be generalized to hearing for the same reasons it fails for colour vision.

I deal with taste and smell together since they are often thought to be interrelated. Tastes and smells each fall into general categories. With taste we have sweet, salty, sour, bitter, and combinations of each in various proportions, and with smell we have the basic categories of minty, floral, ethereal, musky, resinous, foul, acrid, and again, combinations of each.[21] Thus, with each sense, as with colours we have primary elements (like primary colours) which are themselves ordered and can give rise to harmonious or dissonant complexes when combined in various proportions with one another. So although the number of basic phenomenal elements is much lower for taste than it is for vision, they nevertheless share the central relational and structural properties. In the case of smell it appears that the number of basic phenomenal elements is the same as the number of pure colours on the spectrum.

It seems, given the structural similarities between colour qualia and gustatory and olfactory qualia that we can imagine inverting them so that the basic elements are reversed which will also give rise to inversions of complexes of these elements. What I taste as

---

[21] I take these broad categories from Diane Ackerman, A Natural History of the Senses (New York: Vintage Books, 1990) p. 11. She also makes a comparison between these categories and primary colours.

sweet will taste bitter to someone with inverted gustatory qualia, and what I smell as minty will smell acrid to someone with inverted olfactory qualia. Once again, it might initially seem as though the behaviour of those experiencing inverted olfactory and gustatory qualia will be indistinguishable from the behaviour of normal observers, for such people will have learned to call the sensation they experience when eating peanuts "salty" even though they taste sour, and call the smell of roses "floral" even though they smell foul.

While it seems possible for a normal observer and someone with their qualia inverted in the specified manner to agree in this kind of linguistic behaviour, it appears that we once again run into problems when we consider the emotional connections with these qualia. Certain sorts of smells and tastes are very unpleasant. When normal people taste something very bitter the unpleasant character of the taste sensation manifests itself in behaviour. Ordinarily people will make disapproving sounds, will wrinkle their faces, stick out their tongues, and avoid eating any more of the substance in question. Similarly, if one smells something very foul one will cover one's nose and express distress over the perceived odour. As with the corresponding suggestion about pain, it seems that a strong case can be made for the claim that there is an affective component to such sensations, and that this component is partly constitutive of them. For although unpleasant tastes and smells are not painful in the ordinary sense, their extremely unpleasant character, together with the strong behavioural responses they illicit, renders them quite analogous to sensations of pain.

If an individual, call her "Sally," has her gustatory and olfactory qualia inverted, then she will smell things that we call "floral" as foul, and will taste things that we call "sweet" as bitter. This means Sally will do some very strange things. She will not enjoy sniffing sweet smelling flowers or eating chocolate. Instead, she will express the kind of disapproval the rest of us do when we smell rotten eggs or drink sour milk. Conversely, Sally will have a strange capacity to endure those sorts of olfactory and gustatory

experiences the rest of us wish to avoid. These are significant behavioural differences which would suggest to us that something is wrong with Sally's qualia.

To make Sally's behaviour completely indistinguishable from our own we would have to imagine that although Sally really dislikes the taste of chocolate that she will somehow behave as though she enjoys it, and hence will gorge herself on the awful, bitter tasting stuff at any opportunity. We would also have to imagine that although Sally really hates the smell of flowers she will behave as though she loves the foul scent and will happily join us for an afternoon at the botanical gardens. We will also have to imagine that Sally will steadfastly avoid the sorts of gustatory and olfactory experiences that she in fact enjoys. Although the experiences of bitter tastes and foul smells are not painful in the ordinary sense, the inversion hypothesis requires us to accept the idea that an individual could withstand extremely unpleasant sensations without ever expressing the least bit of discomfort, whether in linguistic terms or simply in terms of avoidance behaviour. I think it is clear that such a situation is extremely unlikely, and hence, that the inversion of qualia for these two remaining senses is very implausible.

When considering the sense of touch, temperature appears to be the best candidate for inversion. The phenomenal feel of roughness and smoothness might appear to be analogous to colour qualia in the necessary respects, but it could easily be argued that there is a strong conceptual connection between such phenomenal elements and the nature of the surfaces that cause them which is lacking in the case of colour qualia. Tactile sensations of warmth and coldness seem to be more like colours in this respect than sensations of roughness and smoothness, for it is not obvious to us in the way it is with roughness what it is about an object that produces in us a sensation of heat. Sensations of warmth and cold also represent good candidates for a strong analogy with colours because they organize themselves into a structured continuum in the way colours do (ranging from painful to cold, to warm to painful again) and arguably form complexes (I form such complexes

every time I take a bath by running both hot and cold water). To invert our temperature qualia, then, would, like the inverted spectrum, involve flipping the continuum of phenomenal elements. In this case what I feel as warm the abnormal perceiver feels as cool, and so on.

Again, this inversion seems possible at first glance. We can imagine that Tom's temperature qualia are inverted, but that since he has learned to use words like "warm" and "cold" under the same conditions as the rest of us, his verbal behaviour is no different from ours. Like us, Tom consistently identifies things such as ice cream and winter days as cold, and stove tops and recently prepared porridge as hot. Furthermore, Tom's bodily behaviour might appear to be no different from ours. He quickly withdraws his hand from items we identify as very hot because for him such items feel very cold, and vice versa. Also, he does things like runs his hands under warm water after he has been outside in a snowball fight (since it makes them feel cooler) and splashes his face with cold water on a hot day (since that makes him feel warmer).

All of this remains plausible even when we consider the thermal behaviour of bodies. Objects feel warm or cold depending on the relative mean molecular kinetic energy of the bodies in question. If the kinetic energy of water is higher than that of my hand, then the water will feel warm to me. If it is a great deal lower it will feel cold. Our sensations of heat and cold, then, are primarily determined by the rate of heat loss or heat gain to or from the environment rather than by the actual temperature of the objects involved, which is why good conductors of heat feel cooler than poor ones even when they have the same mean molecular kinetic energy.

When Tom says he feels hot he actually experiences the sensation that you and I describe as a sensation of cold. So as the rate of heat exchange from Tom's body to the environment increases Tom will say that he feels colder and colder. This means that on a very cold day, as Tom feels more and more warm he will pile on more articles of clothing

to make himself feel cooler. However, even though Tom feels warm on a cold day his body temperature will continue to plummet the longer he remains outside. In order to counteract the effect his body's natural defenses will take over in an attempt to raise his body temperature and Tom will start shivering. Thus, Tom's behaviour seems completely indistinguishable from our own.

Given the medical research in support of the idea that all sensations have affective components, and given that we have seen convincing evidence that this is the case for the other sensations sharing the structural features of colour qualia, we have good reason to expect that the same relationships hold for sensations of temperature. Hence we should expect that there are affective differences corresponding to different sensations of temperature (e.g., hot and cold, though perhaps not the extremes since they are both painful), in which case such differences can be used to argue against the inversion of temperature qualia. For if there are standard connections between sensations of warmth and these emotions differ from those connected with sensations of cold, then we would be able to identify emotional differences in the person who has inverted temperature qualia. For given these connections it follows that the affective component of Tom's experiences would be inverted and this is a fact that could be discovered if we designed certain tests to capture Tom's emotional responses to different temperatures.

Such differences are not hard to imagine. For example, compare our willingness to feel hot as opposed to feeling cold. We tend to prefer feeling overheated—up to a point—and typically enjoy lying on a sandy beach in the summer, but quite dislike feeling cold. The enjoyment of the sensation of being nice and warm on the beach is an identifiable, and common, emotional part of the sensation. If Tom's qualia are really inverted, then it seems Tom would quite dislike lying on the beach; instead he would prefer to engage in some activity the rest of us would find disagreeably chilly. Of course this is a crude example. It remains an empirical matter to be sorted out by psychologists exactly what sort of

behavioural differences the emotional ones would lead to, and we might expect that the tests designed to capture emotional differences might reveal quite fine-grained differences that would not otherwise be apparent. Thus, a difference in Tom's temperature qualia will manifest itself in behaviour after all, and so the difference can be functionally captured. It appears, then, that one cannot generalize the inverted spectrum argument to work with the sense of touch.

The conclusion arrived at for sight, hearing, taste and smell, and touch is the same. In each case an undetectable inversion of the associated phenomenal elements is extremely implausible, for such inversions will, because of emotional connections with certain sensations, lead either to behavioural differences, in which case the inversion is detected, or else to unimaginable states of affairs. Admittedly, the claim that the affective component of such sensations is partly constitutive of them, and hence non-detachable, has less than perfect support. However, there is some support for this connection. First, given the plausibility of this hypothesis for pain, and the similarity in structure of the mentioned gustatory and olfactory experiences to pain, this connection appears quite plausible in these cases. Second, views in aesthetics and speech-prosody strongly suggest such connections for auditory qualia. Third, medical research suggests that such affective connections are in fact generalizable to all sensations, including those of temperature. And finally, we have seen that work in environmental psychology supports this connection for sensations of colour. Given the similarities between all of the phenomenal elements that we have considered we have good reason to believe that the inversion hypothesis fails for all senses. In this case functionalism is vindicated and we are left without persuasive reasons to think that qualia are epiphenomenal or pose any special difficulties for physicalism.

2. Jackson's Knowledge Argument

In the previous section I examined the claim that qualia create a serious difficulty for a particular version of physicalism, that being functionalism. Given the inadequacies of the inverted spectrum argument against functionalism there is as yet no reason to suspect that qualia pose any special problem for physicalism as a general hypothesis. However, there are a number of philosophers who believe that qualia do represent a basic challenge to physicalism, regardless of which version of physicalism one adopts. One such philosopher is Frank Jackson. In his controversial paper "Epiphenomenal Qualia," Jackson claims to show that there are facts the physical sciences cannot capture and hence that there are non-physical facts. Furthermore, Jackson argues that the non-physical properties to which these facts refer are epiphenomenal in the traditional sense of being causally impotent. By showing that there are epiphenomenal non-physical properties, Jackson believes he demonstrates the falsity of physicalism. My aim in this section is to assess the cogency of Jackson's argument. My investigation is divided into three parts. First, I explain Jackson's intuition about subjective experiences and sketch out his "knowledge argument" against physicalism. Second, I examine two versions of the "no information" reply frequently offered in response to the argument and show they are unconvincing. Finally, I show that an alternative line of objection is more promising than the first, but that it does not require us to reject Jackson's argument completely; instead, I show that all it requires is a weakening of Jackson's conclusions. While the knowledge argument does not entail the existence of non-physical properties as Jackson claims, it nevertheless seems to entail the denial of the explanatory completeness of physicalism.

Jackson is a self-proclaimed "qualia freak." He takes the existence of qualia very seriously, and, like many who have a fondness for qualia, thinks they represent a special obstacle to physicalism. By "qualia" Jackson means to refer to the qualitative characteristics of conscious mental states. These include such things as, for example "the hurtfulness of pains, the itchiness of itches, pangs of jealousy,"[22] and so on. The obstacle they create for physicalism is that one can know everything physical there is to know about human beings, but, it seems, one does not thereby learn about the hurtfulness of pains or the itchiness of itches. This suggests that qualia are non-physical properties.

Jackson's intuition was also held long ago by Leibniz. In Section 7 of the Monadology Leibniz says,

> We are moreover obliged to confess that perception and that which depends on it cannot be explained mechanically, that is to say by figures and motions. Suppose that there were a machine so constructed as to produce thought, feeling, and perception, we could imagine it increased in size while retaining the same proportions, so that one could enter as one might a mill. On going inside we should only see the parts impinging upon one another; we should not see anything which would explain a perception. The explanation of perception must therefore be sought in a simple substance, and not in a compound or in a machine.[23]

Although the point is put somewhat differently by Leibniz, the intuition is the same as Jackson's (and entailed by Locke's inverted qualia hypothesis). Both authors share a belief that the physical information made available by examining the mechanical functioning of the human organism is in some way inadequate when it comes to explaining certain features of perceptual experience. And, like Leibniz, Jackson draws an inference from this explanatory failure to the existence of non-physical items. Let us now turn to the argument Jackson offers in support of this intuition and his inference to the existence of non-physical properties.

---

[22] Jackson, "Epiphenomenal Qualia," p. 127.

[23] Gottfried Leibniz, "Monadology," in Leibniz: Philosophical Writings, ed. G.H.R. Parkinson, trans. Mary Morris and G.H.R. Parkinson (London: J.M. Dent and Sons Ltd, 1973) p. 181.

Although Jackson provides two versions of the knowledge argument it is the second of these that has attracted the most attention. I will therefore limit my discussion to this version. Jackson asks us to imagine a brilliant neurologist, Mary, who has been imprisoned in a black and white room her entire life and who learns about the world through black and white television monitors. Although she has never experienced colours herself, through many years of study Mary learns not just a lot, but everything physical there is to know about the neurophysiology of colour vision. Jackson asks,

> What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and physicalism is false.[24]

So, since by hypothesis Mary has all the physical information there is to have about colour vision yet learns something when she leaves the room, there is more to know than the physical information. This entails that there is some fact (what colour sensations are like) that is a non-physical fact, and since there are non-physical facts physicalism is false.

Jackson is careful to distinguish this argument from a similar one against reductionism developed by Thomas Nagel in "What Is It Like to Be a Bat?"[25] In this article Nagel argues that there is no amount of physical information that could tell us what it is like to be a bat, or to have the point of view on the world possessed by a being sufficiently unlike us. Jackson's knowledge argument differs from Nagel's in that it is not concerned with what it would be like for Mary to be someone else who has colour experiences. The issue is whether or not she learns anything about others when she leaves her room. According to Jackson she does since she learns what it's like to see colours.

---

[24] Jackson, "Epiphenomenal Qualia," p. 130.

[25] Thomas Nagel, "What Is It Like to Be a Bat?," Philosophical Review 83 (1974) pp. 435-450.

In the existing literature there are three standard replies to the knowledge argument. The first two of these can be characterized as different versions of what is sometimes called the "no information" reply. The central aim of the no information reply is to show that Mary does not acquire new information or learn new facts when she leaves her black and white environment, but instead gains something else. By denying that Mary acquires new information the critics hope to show that Jackson's conclusion doesn't follow from the knowledge argument.

According to the first version of this reply what Mary gains when she leaves her room is not information but is rather an ability; it is knowledge-how rather than knowledge-that. The ability in question is often thought to be an ability to imagine. Whereas Mary could not previously imagine what it's like to experience red, once she leaves her room she comes to possess a new imaginative ability: she can imagine seeing red. This response has been articulated most explicitly by Laurence Nemirow, but has also been endorsed by David Lewis and taken up in a slightly different form by Paul Churchland. Nemirow formulates the reply as follows:

> Knowing what it's like may be identified with knowing how to imagine.
> The more seriously we take this ability equation, the easier it becomes to resist the knowledge argument. The latter assumes that science cannot convey what it's like to see red. The premise is uncontentious, for science does not seek to instill imaginative abilities. But the knowledge argument concludes that physical science cannot describe certain information about seeing red. The inference is invalid because it presumes that knowing what it's like is propositional knowledge rather than an ability.[26]

According to Nemirow, then, what Mary gains is an ability to imagine seeing red. If we can identify knowing what it's like to see red with the ability to imagine seeing red, Jackson's conclusion doesn't follow because on this account Mary learns no new facts.

---

[26] Laurence Nemirow, "Physicalism and the Cognitive Role of Acquaintance," p. 493.

Is there any reason to make this identification, aside from the fact that it apparently dismantles the knowledge argument? Nemirow thinks there is. He points out that communicating an ability one has to others who lack it is extremely difficult (to the point where we might say that abilities are almost inexpressible) which gives us further reason to suspect that what is involved in knowing what it's like to see red is an ability. For this suggests that the inexpressibility typically associated with knowledge of qualia (for instance, when people say things like "I can't tell you what vegemite tastes like; you have to taste it for yourself") might not be derived from the kind of information involved, as "qualia freaks" frequently assume, but is merely the sort of inexpressibility common to all abilities. Those who claim that qualia have the peculiar feature of being incommunicable have confused the source of this inexpressibility and mistakenly assumed that qualia convey genuine information. Thus, Nemirow's proposal to identify "knowing what it's like to X" and "being able to imagine X-ing" has the advantage of explaining the apparent incommunicability of qualia without invoking the idea that qualia convey information.

Jackson vehemently resists this version of the no information reply. In a defense of his original paper he says,

> The knowledge argument does not rest on the dubious claim that logically you cannot imagine what sensing red is like unless you have sensed red. Powers of imagination are not to the point. The contention about Mary is not that, despite her fantastic grasp of neurophysiology and everything else physical, she could not imagine what it is like to sense red; it is that, as a matter of fact, she would not know. But if physicalism is true, she would know; and no great powers of imagination would be called for. Imagination is a faculty that those who lack knowledge need to fall back on.[27]

This hardly seems a satisfactory reply to Nemirow's response to the knowledge argument. In the first place, Jackson's characterization of the imagination is obviously too narrow to be plausible. I know what it's like to see red and I can imagine seeing red, so my

---

[27] Frank Jackson, "What Mary Didn't Know," Journal of Philosophy 83 (1986) p. 293.

faculty of imagination can hardly be characterized as something I rely on only when I lack knowledge. Thus, Jackson is not taking the identification suggested by Nemirow seriously, but instead dismisses it out of hand by simply assuming that imagining red and knowing what it's like to see red are different. Second, Jackson's insistence that Mary simply "would not know" what seeing red is like does nothing to repudiate the claim that the kind of knowledge Mary acquires is not propositional knowledge, but a kind of know-how. Third, and most importantly, Jackson has not given us reasons for thinking that the identification of "knowing what it's like to see red" with "knowing how to imagine seeing red" is "dubious" as he claims. Jackson may be correct about this, but without further argument there is little reason to agree with him. What we need to do is determine whether or not this identification is in fact "dubious" as Jackson assumes.

The place to begin in assessing the plausibility of the identification required by this version of the "no information" reply is to ask whether the account of the imagination involved is a credible one. In order to do this we do not need to develop anything as complex as a full-blown "theory of the imagination"; all we need to do is ask what Nemirow's account of the imagination requires and then determine whether it is true to experience. In order to identify "knowing what it's like to X" and "being able to imagine X-ing" we need to assume that the objects of imagination must themselves be qualitatively similar to real experiences. That is, when one imagines phenomenal redness whatever it is that one holds "before the mind's eye" must itself resemble the perception of something red. Similarly, whenever one imagines having a toothache whatever it is that one is imagining must resemble a painful experience.

Nemirow, following Berkeley and Hume, adopts precisely this account of the imagination. On this view to imagine something is to represent some particular rather than to entertain something like a universal:

> To visualize red, for example, is to apprehend neither the quality of being red nor the quality of seeing red; it is only to represent particular perceptions of a particular shade of red. Similarly, to imagine pain is not intellectually to apprehend the quality of being in pain; it is to represent a particular painful experience.[28]

Although one can "represent" particulars in any of a number of ways, the form of representation suggested by Nemirow must involve reflection on a particular quale. For how else could one do justice to the "particular perception of a particular shade of red"? Without the particular quale being present in my mind it seems as though I would merely be thinking of redness or some sort of general concept. Nemirow goes on to say that "Berkeley and Hume dispelled the philosophical clouds surrounding imaginative representation."[29] I think there is good reason to doubt they did, in which case Nemirow is simply adding to the confusion. I can make no claim about the imaginative abilities of others, but when I imagine the Canadian flag or a toothache I do not have an experience resembling the perceptions of anything red or painful (try it and see for yourself). My imagined flag possesses nothing like the salient characteristics of a perceived flag, or even of something like an after-image. Similarly, my imagined pain is not itself painful. Instead, I find myself thinking of pain-behaviour which can hardly be identified with a pain quale. I suspect that what I am doing when I imagine such things is precisely what Nemirow denies: I am thinking about the concepts of redness and pain.[30]

This is not to deny that one might find oneself presented with something faintly resembling phenomenal redness (perhaps "less vivid" than the original) when one imagines something red. Indeed, some people may have more active imaginations than I do. But when it comes to imagining the subjective aspects of other bodily states it seems less

---

[28] Nemirow, p. 495.

[29] Ibid.

[30] For a more detailed discussion of such worries about the imagination see Daniel Dennett, Consciousness Explained (Boston: Little, Brown and Company, 1991) pp. 55-65.

plausible that such imaginings qualitatively resemble their real counterparts as Nemirow's account requires. Imagined pains do not qualitatively resemble real pains, and so are not themselves hurtful. If they were I wouldn't like them, but I am as a matter of fact quite indifferent to imagined pains.

I suspect that Nemirow has gone wrong by confusing the act of imagination with the more specific act of imaging. As I mentioned, one can imagine or represent an object in any of a number of ways. One way Mary might imagine seeing red is by imagining that she has the ability to discriminate between ripe and unripe tomatoes without the aid of her colour-detecting instruments. It is therefore possible for Mary to imagine seeing red without ever being presented with a red quale, in which case she still would not know what it is like to see red. Given this, the two acts can hardly be identified as Nemirow suggests. The act of imaging, however, requires prior acquaintance with a particular quale. To image red is to be able to form a mental image of red. This is not possible for Mary until she herself is appropriately acquainted with a red quale, either by leaving her black and white environment, or by somehow entering the brain state people are ordinarily in when they experience red. Nemirow seems to have assumed, incorrectly, that to imagine red is to be able to image red and that one can image red without ever having had an experience of red.

The point of these observations is that since imagined pains are not hurtful and other imaginative states do not qualitatively resemble actual perceptual experiences there is little reason to identify qualia with imaginative abilities. The conclusion we can draw from this insight is that it is unlikely we can explain Mary's new knowledge in terms of an ability to imagine. Once Mary knows what it's like to see red she does in fact acquire a series of new abilities, including the ability to image red, but these abilities are not identical to her knowing what it's like to see red, they are consequences of having this knowledge. This version of the no information reply, therefore, does not dismantle Jackson's argument.

The second version of the no information reply was first formulated by Terence Horgan and has also been proposed by Paul Churchland. Rather than rely on a suspicious identification of abilities as Nemirow does, they instead identify what they think is a subtle fallacy in Jackson's argument. Horgan and Churchland claim Jackson equivocates on "knows about" in his thought-experiment. When Jackson says that Mary "knows" all the physical information and "knows" what it's like to see red, he does not use the word "knows" univocally. Churchland summarizes the problem this way:

> In short, the difference between a person who knows all about the visual cortex but has never enjoyed a sensation of red, and a person who knows no neuroscience but knows well the sensation of red, may reside not in what is respectively known by each (brain states by the former, qualia by the latter), but rather in the different type of knowledge each has of exactly the same thing. The difference is in the manner of the knowing, not in the nature of the thing(s) known.[31]

So according to Churchland, what Mary knew before she left her black and white world was the same thing she came to know when she left her room, she just came to know it in a different way. Before she escaped Mary knew the experience of red by description and now she knows it by acquaintance. The difference Jackson emphasizes between Mary's knowledge of the sensation of red on these two occasions can be accounted for in the manner of the knowing instead of the thing known. If this is correct, then Jackson is mistaken in drawing the inference from the knowledge argument that Mary comes to know a non-physical fact or property.

Jackson has an interesting rejoinder to Churchland's argument. He claims that the distinction between knowing something by description and knowing something by acquaintance is irrelevant to the argument. Let's assume that the sensation of red is identical with some brain state, call it ϕ, and that Mary knew that when others said things like "I see red" (and it was true), they were in that brain state. When Mary enters the coloured world

---

[31] Paul Churchland, "Reduction, Qualia, and the Direct Introspection of Brain States," Journal of Philosophy, 82 (1985) p. 24.

the point is not that she learns something new about herself, for we are assuming she never was in ϕ when she was in her room, so there was no such fact for her to learn about herself. What is significant is that Mary learns a new fact about others. She was in the position to observe other people in ϕ before her escape, but now that she sees colours herself she comes to know more about ϕ than she did before; she learns what it's like to be in ϕ and this is a fact about others that had previously escaped her.

Earl Conee develops this rejoinder to the Horgan-Churchland reply further and puts it as follows:

> Now consider what happens when Mary first sees something red. She becomes aware of a simple visual presentation—the look of something red. It can be maintained that she is then aware of the physical property which is phenomenal redness [what I have described as ϕ], something that she was already acquainted with when she learned the physics of colour-perception. But it seems beyond doubt that something new is also involved in her experience. Things do not seem the same to Mary as they did when she was aware of phenomenal redness by means of the representation consisting in electrochemical notions. Only some new element can account for this difference in how things appear.[32]

This seems correct. There is a new property involved in Mary's knowledge of redness and this property was not one she could have known before.

There is a further problem with the Horgan-Churchland reply. In speaking of Mary coming to know the same thing in different ways they make the questionable assumption that Mary's redness quale can appear to Mary in two different ways: either as a specific brain state (from the standpoint of neurology), or as phenomenal redness (from the first-person perspective).[33] There is, however, an oddity about speaking this way, for it assumes that a quale is something that can itself be an object of experience and there is

---

[32] Earl Conee, "Physicalism and Phenomenal Properties," Philosophical Quarterly, 35 (1985) p. 300.

[33] Unlike the propositional attitudes, Churchland thinks that the qualitative states associated with certain experiences will not be eliminated along with folk psychology, but will be identified with certain physical states. Hence he can continue to talk about qualia without undercutting his own eliminativist project. For some indication of this see Churchland, Matter and Consciousness, (Cambridge, Mass: MIT Press, 1984) p. 40.

good reason to resist this assumption. On the standard interpretation Mary's redness quale is the way a red object appears to her under normal conditions, it is not itself an object of experience, for how could appearances themselves appear to perceivers in different ways? To claim that they do is a confusing and misleading way of speaking. Since the Horgan-Churchland version of the no information reply requires this way of speaking about qualia, it appears misguided.[34]

The third and final objection to Jackson I want to consider is proposed by Conee. While Conee thinks the Horgan-Churchland response is a poor one, he does not go on to draw the conclusion that Mary learns a non-physical fact. To admit that the property of which Mary becomes aware is a non-physical property is to isolate it from the causal relations that constitute the physical processes involved in colour vision, which means that the property is epiphenomenal. For if the property were non-physical but causally efficacious, then Mary would have noticed mysterious gaps in her causal explanation of colour vision. Since the assumption is that she notices no such gaps, Jackson thinks qualia must be epiphenomenal. According to Conee this is too implausible to be true so we ought to assume the property is causally efficacious. Besides, if qualia were epiphenomenal in the sense implied by Jackson, it is unclear how anyone could know or speak about them.

Conee claims that Jackson has given us no decisive reasons for thinking that phenomenal properties are epiphenomenal "except for Jackson's argument for the non-physical nature of the qualities, together with the difficulties concerning non-physical interventions."[35] Since there is no conclusive reason to deny that phenomenal properties are causally efficacious, perhaps we can reconcile Mary's inability to know such properties with their causal efficacy. The benefits of such an analysis are that we retain a form of

---

[34] For a similar objection see Paul Raymont, "Tye's Criticism of the Knowledge Argument," Dialogue XXXIV, (1995) p. 718.

[35] Conee, "Physicalism and Phenomenal Properties," p. 301.

physicalism and the view that phenomenal properties have causal powers. The solution is to deny that Mary did know everything physical there is to know about colour vision when she was in her black and white room. According to Conee, then, we need to adjust the conclusions of the thought-experiment and claim instead that phenomenal redness is a physical property but one Mary couldn't learn about in her studies. On this view one cannot learn everything physical there is to know about colour vision in a black and white environment. We can explain the apparent completeness of Mary's knowledge of the causal explanation of colour vision (despite the missing efficacious property) by the possibility that such properties play an intermediate causal role at some stage in the physical processes involved in colour vision which could pass unnoticed without introducing apparent gaps in the causal story. For instance, physical property P may be both necessary and sufficient for the phenomenal property Q, which is in turn necessary and sufficient for physical property R. Given this the sequence might appear such that P is necessary and sufficient for R without ever mentioning the causal role of Q. Conee's conclusion is as follows:

> If we hold on to what seems plausible—Mary's discovery of the phenomenal quality, the causal role of such qualities, and the physical character of whatever has such a role—then it is worthwhile to exploit this possibility by supposing that phenomenal qualities are physical, properties that have the causal traits of Q [i.e., they are intermediate causes].[36]

Unfortunately, it should be obvious that we cannot place too much weight on the possibility that phenomenal properties are physical yet always play an intermediate role in causal relations such that they are continually overlooked by scientific accounts of experience. For it seems odd that it is just these sorts of properties (for the same must be true of other phenomenal properties) that are so adept at playing hide-and-seek with our scientific accounts of the world. We cannot avoid the question, which Conee apparently thinks he can, of why such properties are so elusive from the standpoint of the physical

---

[36] Ibid., p. 302.

sciences. However, I think that Conee is on the right track when it comes to evaluating Jackson's argument. It is, as he says, possible that the property Mary comes to know is a physical property. All we need to make this claim plausible is an account of how phenomenal redness can be a physical property, yet escape scientific description. For once we have shown how it is possible for phenomenal redness to be a physical property in a way that is consistent with the main intuitions of Jackson's thought experiment, the burden of proof lies with Jackson to show us why the property cannot be physical.

There is a suggestion in Conee's discussion about how we might answer this question. Phenomenal properties might be physical properties of brain states that can be known only by being in the requisite brain state. This is not surprising since what seems crucial to knowing phenomenal redness is being in the appropriate brain state oneself. For even if Mary never left her black and white environment she could still come to know what it's like to see red if we artificially stimulated her brain in the proper way. Hence, it is tempting to identify "Knowing what it's like to X" with "Being in brain state Y." We need to be careful about what this commits us to, however. The phenomenal property cannot, as Churchland suggests, simply be the way one's brain state appears to oneself as opposed to others, for this assumes that one's brain states are objects of experience in the way that ordinary objects are. While it is true that I can study my own brain states in the same way as a neurologist might (from the "outside," as it were), I am doing something very different when I reflect upon the character of my experiences from my ordinary perspective. Without adopting the standpoint of the neurologist it makes little sense to speak of the way my brain states appear to me because my brain states are the very processes of thought and perception themselves, not the objects of thought and perception.

These observations provide some headway to understanding why qualia escape capture in physicalist terms but can be made more plausible with the following considerations. The interesting thing about phenomenal properties is that, despite Jackson's

claims to the contrary, they appear to play a causal role in our behaviour. Phenomenal properties ordinarily have the role of enabling us to make certain discriminations. For it is on the basis of my colour qualia that I am able to distinguish ripe from unripe tomatoes, red from green, and so on.[37] When we consider their usual discriminative role a clue to how such properties might escape scientific capture can be found by examining the phenomenon of blindsight.

Cases of blindsight are well documented and beloved by philosophers of mind. A person who experiences blindsight claims to have no visual awareness at all in part of their visual field yet can "guess" far better than average whether an object is present in that part of the field, what shape it is, what colour, and so on. What appears to be going on is that such people can "see" the object in that area yet have no awareness of what they are seeing. It is as though the perception of the object is unconscious. Weiskrantz, who coined the term "blindsight" defines it as follows: "visual capacity in a field defect in the absence of acknowledged awareness."[38] Here we have an interesting case where the phenomenal property is not detectable either by the neurologist (if Conee is correct) or by the individual in the brain state approximating the one ordinarily associated with the phenomenal property (I say "approximating" because the assumption is that the brain has suffered some physical damage). The interesting thing to note, however, is that cases of blindsight require prior phenomenal discrimination. There are no cases where someone blind from birth can

---

[37] William Seager calls the concepts involved in making the discriminations associated with how things look or taste, etc., "substantial" concepts, and also claims, as I do, that one must have the corresponding experience in order to come to possess the concept. See William Seager, "Critical Notice of Fred Dretske, Naturalizing the Mind," Canadian Journal of Philosophy 27 (March, 1997) pp. 89-93.

[38] L. Weiskrantz, Blindsight: A Case Study and Implications (Oxford: Clarendon Press, 1986) p. 166.

suddenly discriminate colours yet sincerely denies having any awareness of colour.[39] Were this to happen it would surely be very surprising. It is plausible, then, to suggest that in cases of blindsight the phenomenal property is part of the causal complex of blindsight (otherwise, the displayed discriminations would not have occurred because the subject could not have acquired the necessary colour concepts or connect colour words to objects in the world), yet seemingly plays no active or direct causal role. If the phenomenal property did play an active role, then the person in question would explicitly rely on his or her quale and have an awareness of doing so. Thus, I suggest that phenomenal colour experiences play a special causal role: they activate the capacity for colour discrimination. But this does not mean that the causal role remains active. It is not active in cases of blindsight and therefore need not be active in other cases in which we make colour discriminations. The phenomenal property is therefore causally efficacious (it is part of the causal complex of vision and blindsight) but is not obviously involved in the causal mechanisms of colour vision.[40] By linking phenomenal properties to discriminative abilities in this way we can see how such properties can be characterized as physical properties that play a causal role, yet might be elusive by ordinary standards of investigation. This characterization of phenomenal properties as physical properties connects with the conclusion drawn in the first half of this chapter, that the inverted spectrum does not constitute a genuine possibility. For assuming that human brains are similar this identification of phenomenal properties with physical properties of the brain

---

[39] Although Weiskrantz does not consider colour discriminations in blind fields, he does mention another study that does. See A. Damasio et. al., "Nervous Function After Right Hemispherectomy," Neurol. 25 (1975) pp. 89-93.

[40] My thanks to Evan Simpson for this intriguing suggestion.

suggests that if the physical properties are similar that the phenomenal properties should be similar also.[41]

It might seem that the claim that phenomenal properties can be known only by entering the appropriate brain states entails that such states are essentially private. The notion that qualia are private is all too common and leads to a number of philosophical puzzles. Fortunately, it doesn't follow from what I have said that phenomenal properties are private in any problematic sense, in which case the usual philosophical quandaries arising from the notion of privacy are avoided. The only sense in which such mental states are private is that one cannot access them by investigating the physical functioning of the brain and central nervous system. This does not mean that such states are essentially private. As I said, Mary can learn about the phenomenal redness known by others if she enters the required brain state herself. There is therefore no principled obstacle preventing Mary from gaining knowledge of other people's mental states and certain of their properties.

Such an account of phenomenal redness does not completely dismantle Jackson's argument. However, it does force us to weaken the conclusion we can draw from it. It follows from Jackson's argument that qualia cannot be captured by the terms of a physical theory, provided we think of physical theories as accounting for the nature of the physical world independently of particular observers. However, we cannot go on to draw the further conclusion that there are non-physical properties, for we have seen a way in which we can speak of properties such as phenomenal redness as physical properties while denying that they can be known through scientific observation (in the normal sense). Neither can we conclude that such properties are epiphenomenal in the traditional sense.

---

[41] In the next chapter I will spell this out in more detail with the claim that qualia supervene on physical states.

For we have seen no compelling reasons to deny that phenomenal properties have causal efficacy. Instead, we must stop with the weaker conclusion that physicalism is explanatorily incomplete since there are facts, such as facts concerning what it is like to perceive a red object, that escape its capture. Thus, the challenge posed by qualia to physicalism should not be regarded as a challenge to physicalism's ontological claims at all. Instead, qualia identify a limit to the explanatory power of physicalism. Of course the important question remains whether this explanatory failure on the part of physicalism entails the denial of physicalism's claim to explanatory completeness. While the answer might appear to be an affirmative one, I suggest in the next chapter that in fact this is not the case.

Chapter 4

Supervenience

The groups of objections I examined in the previous two chapters focused on two disparate kinds of mental phenomena. My discussion of Davidson's anomalous monism in Chapter Two had as its point of focus the propositional attitudes, whereas my discussion of functionalism and Jackson's "knowledge argument" in Chapter Three dealt with the qualitative states of consciousness.[1] Before I proceed any further it would be useful to draw some general conclusions from these two discussions and, where possible, bring them together and appreciate where they intersect.

The conclusion I arrived at in my discussion of the epiphenomenalist challenge to anomalous monism was that although the mental states or events with propositional content referred to in every day psychological explanations of human behaviour have causal powers, the identification of such states does not causally explain behaviour. Since such states explain behaviour nevertheless, it seemed that they must explain behaviour in a different way than causal explanations do. Working within a broadly Davidsonian framework, I concluded that we explain intentional actions by showing how they are rational in the light of an agent's other mental states, and that situating actions within a context of rational behaviour in this way ordinarily serves us quite well as an explanation

---

[1] Little, if anything, hangs on the distinction between the propositional attitudes and qualia. I invoke the distinction simply for convenience since it is one that is generally accepted. In Chapter 5 I briefly outline a proposal for how we might do away with this distinction by following suggestions made by Evan Simpson in "Sensation Deconstructed," in Entities and Individuation: Studies in Ontology and Language in Honour of Neil Wilson, ed. Donald Stewart (New York: Edwin Mellin Press, 1989) pp. 153-164.

provided, of course, that the reasons identified also cause the actions in question. This leads to a distinction between two species of explanation (causal explanations and reason explanations) which seems to raise a problem for physicalism. Since reason explanations cannot be causally construed it seems they cannot be incorporated into physicalist accounts of behaviour, for there is no way to connect such explanations to the causal explanations dealing in physical predicates. Given this it seems to follow that reason explanations are autonomous, in which case anomalous monism (as an articulation of physicalism) is explanatorily incomplete.

My conclusion in the discussion of phenomenal properties was similar. Although it was possible to identify phenomenal properties with physical properties of the brain and bestow upon them causal powers (just as it is for mental states with propositional content), it seems as though something is left out of the understanding of perceptual experience provided by physicalist accounts of perception. Such accounts can never capture the subjective aspects of such properties, the "what it is like to . . ." have certain experiences. This means that physicalism necessarily "leaves something out" of its account of experience and hence appears once again to be incomplete at the explanatory level, for it leaves certain facts unexplained or uncaptured.

The sort of incompleteness identified in the last two chapters is therefore similar. Both sets of properties reveal that physicalism suffers from certain explanatory inadequacies, and it seems to follow from these inadequacies that physicalism is incomplete at the explanatory level. While this is surely a less damaging blow to physicalism than the negation of its ontological thesis would be, it nevertheless constitutes a serious objection, especially when much of the job of a physicalist theory of mind is to account for and to explain the details of human behaviour and experience. However, I have indicated that these objections might not be as compelling as they initially seem. For it does not necessarily follow from the claim that the forms of physicalism I have discussed suffer

from certain explanatory inadequacies that physicalism is thereby incomplete at the explanatory level, for it might be the case that reason explanations and facts about qualia depend on physical explanations and physical facts. If this is the case, then the epiphenomenalist conclusion does not follow, for it turns out that qualia and reason explanations are not autonomous after all. This defense of physicalism rests upon the property known as "supervenience."

Supervenience, as a philosophical concept, was first developed in moral philosophy. Although R. M. Hare denies it, he is probably the first to use the term "supervenience" in the modern philosophical sense explored in this chapter.[2] The relation denoted by this term, however, goes back farther than Hare's philosophy and can be found, for instance, in the moral philosophy of G. E. Moore.[3] Hare (and, implicitly, Moore) claimed that moral properties such as goodness and badness supervene on non-moral or descriptive properties. The concept of supervenience, then, was construed by them as a relation holding between two families of properties or predicates: the "supervenient" family and the "base" family. In this case whether one is described as "good" or "bad" (the supervenient family) supervenes on what actions one performs (the base family).

This relation was thought to capture several significant ideas. First, Hare and Moore both believed that the relation is one of dependence and determination. What moral properties one has (e.g., whether one is good or bad) is dependent on and determined by the descriptive properties true of that person. Second, supervenience constrains the distribution of the related properties in interesting ways. For instance, if the moral supervenes on the descriptive, then two people who share all of their descriptive properties

---

[2] R. M. Hare, The Language of Morals, (London: Oxford University Press, 1952).
[3] G. E. Moore, Philosophical Studies, (London: Oxford University Press, 1922).

(they perform all the same actions under the same conditions, for example) cannot differ with respect to their moral properties (they must both be good or bad). Finally, the relation was thought to be weak enough to deny that the supervenient properties can be reduced to their base properties. Both Moore and Hare resisted the idea that the supervenient moral properties could be analyzed into or identified with their descriptive base properties. After all, while they claimed that two people who share the same descriptive properties are morally equivalent, the converse does not necessarily hold.

Davidson is generally acknowledged as the first to make use of this relation in the philosophy of mind. He follows Hare's example by characterizing the relation in terms of dependence and determination, and likewise denies that the supervenient properties can be reduced to their base properties:

> Although the position I describe denies there are psycho-physical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect. Dependence or supervenience of this kind does not entail reducibility through law or definition . . . .[4]

Since Davidson's initial characterization of psycho-physical supervenience there has been a virtual explosion of literature on the subject. Much of this work is concerned with clarifying the force of "cannot" in Davidson's claim that "there cannot be two events alike in all physical respects but differing in some mental respect . . . ." Hence there are numerous discussions about the proper modal force that should be assigned to psycho-physical supervenience. As a result of the uncertainty about this several modal variants of the relation have emerged along with accounts of the connections between each of these.

---

[4] Donald Davidson, "Mental Events," in Davidson, Essays on Actions and Events (Oxford: Oxford University Press, 1980) p. 214.

The central point of this chapter is not primarily concerned with these debates. Instead, I would like to focus on whether or not the relation of supervenience actually captures the sense of dependence it was thought to. The general strategy in this chapter is to show first, that supervenience can be construed in a way that expresses dependence, and then to show that given this the supervenience of reason explanations and facts about qualia on causal explanations and physical facts defeats the epiphenomenalist objections.

1. Supervenience and Dependence

Debates about supervenience have cooled off over the past few years. Those that remain are focused either on technical points concerning the modal force of, or connections between, different formulations of the relation, or on issues of reduction. Strangely enough the more interesting question—at least for its application to the philosophy of mind—has received little attention. The question is whether or not psycho-physical supervenience expresses a relation of dependence. Of course this will depend largely on how supervenience is understood and formulated. While the concept of supervenience captures the idea of dependence (after all, that is what it was introduced to do), it is not clear that its existing formulations are adequate in this respect. If supervenience is to do any real philosophical work for us, then it is important to determine whether or not it can be rigorously formulated and understood in a way that clearly captures the idea of dependence, for otherwise it takes on the appearance of an empty concept that only seems to serve a philosophical function, when it fact it serves none. Jaegwon Kim, more than anyone else, has shown that the standard formulations of supervenience fail to capture the idea of psycho-physical dependence they were taken to express. While I think Kim is essentially correct about this, I wish to question his rejection of weak forms of supervenience such as Davidson's as candidates for the expression of psycho-physical dependence. Drawing on some themes

from Chapter Two I argue that Kim's failure to appreciate the difference between conceiving of the relation as one that holds between properties and one between predicates re-opens the possibility that weak Davidsonian supervenience is a relation of dependence.

The claim that psycho-physical supervenience expresses the dependence of the mental on the physical was, according to Kim, made for the first time in Davidson's characterization of supervenience in "Mental Events"[5] In two seminal works Kim examines whether the property covariation expressed by psycho-physical supervenience can be seen to include or entail a dependency relation as Davidson claims it does.[6] However, Kim does not limit his investigation to Davidson's formulation of the relation. Instead, he offers three characterizations of his own which have since become the recognized standards, one of which, he claims, is equivalent to Davidson's formulation. They are: "weak," "strong," and "global" supervenience:

> 1. A weakly supervenes on B if and only if necessarily for any property F in A, if an object 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.[7]
>
> 2. A strongly supervenes on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and necessarily if any y has G, it has F.[8]
>
> 3. A globally supervenes on B just in case worlds that are indiscernible with respect to B ("B-indiscernible," for short) are also A-indiscernible.[9]

The difference between weak and strong supervenience is that there is a second modal operator at work in the latter. Weak supervenience guarantees the relation specified between the two classes of properties holds within one possible world only, whereas the two modal

---

[5] Reprinted in Davidson, Essays on Actions and Events (Oxford: Oxford University Press, 1980) p. 214. First published in 1970.

[6] Jaegwon Kim, "Concepts of Supervenience" and "Supervenience as a Philosophical Concept," in Kim, Supervenience and Mind (Cambridge: Cambridge University Press, 1993).

[7] Jaegwon Kim, "Concepts of Supervenience," p. 64.

[8] Ibid., p. 65.

[9] Ibid., p. 68.

operators in the characterization of strong supervenience guarantee the relation holds across possible worlds. To use Kim's example, if being good weakly supervenes on being courageous, benevolent, and honest, then although every person in this world who exemplifies these three properties is necessarily good, there may be some other possible world in which some such person is evil.[10] If goodness strongly supervenes on these base properties, then anyone who possesses them must be good in any world. Finally, global supervenience speaks of the relation as holding generally between "worlds." For example, it claims that if the moral globally supervenes on the descriptive, then worlds that are descriptively indiscernible are morally indiscernible.[11]

While the question about the appropriate grade of necessity that psycho-physical supervenience ought to express might appear esoteric to some, the point is actually quite important. Depending on which grade of supervenience we accept the form of physicalism it expresses is correspondingly stronger or weaker, and there might be independent reasons for preferring one to the other. For instance, those who think that inverted qualia represent a genuine possibility will tend to prefer a very weak form of supervenience since stronger versions don't allow for the possibility of qualia inversion. Alternatively, one might prefer strong supervenience because it might (with some finessing) be seen as entailing that mental properties are reducible to physical properties.[12] Obviously, forms of reductionism represent a much stronger version of physicalism.

Although the comparison is already implicit in Kim's definitions, we might contrast global supervenience with "local" supervenience by characterizing the latter this way:

---

[10] Ibid., pp. 58-60.

[11] It bears observing here that Kim chooses a poor example to express the supervenience of moral on descriptive properties since properties such as courage, benevolence, and honesty are arguably evaluative rather than descriptive.

[12] For such a view see Kim, "Concepts of Supervenience," in Kim.

> A locally supervenes on B just in case individuals that are indiscernible with respect to B ("B-indiscernible," for short) are also A-indiscernible.

Local supervenience can be either strong or weak depending on our intuitions about the appropriate grade of necessity involved in such a claim.

The distinction between local and global supervenience is in part motivated by a division in the philosophy of mind about what elements ought to be included in the physical ground of one's psychological states. On the one hand there are those who advocate what is called "individualism." On this view one's psychological states depend for their identity only on one's intrinsic characteristics, such as one's neurological states. The opposing view, "anti-individualism," is generally connected with a view called "externalism," which claims that much more figures in the identity of one's mental states than this; there are also various relational properties in the physical base which determine the content of one's mental states.

Externalists like Putnam have argued that features of the external world are in part constitutive of one's mental states.[13] He shows this with his Twin Earth argument. Twin Earth is exactly like Earth, except that the stuff that looks and behaves like water on Twin Earth is made of some other substance, XYZ. Since, for Putnam, my belief about water and my doppleganger's belief about water differ in content by virtue of the chemical constitutions of those items (even though, by hypothesis we do not know what water is made of in each world), it follows that what makes our beliefs the beliefs they are goes beyond what happens in our brains. This means that the environment plays a crucial role in individuating mental states. If one's mental states supervene on more than one's brain states, then local supervenience will not suffice. Such forms of externalism therefore

---

[13] For a sampling of this view and the arguments in its favour see Hilary Putnam, "Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind," Journal of Philosophy XCI (1994) pp. 445-517; Representation and Reality (Cambridge: MIT Press, 1988).

demand that the mental globally (or something intermediate between locally and globally) supervenes on the physical.[14]

According to Kim neither weak nor global supervenience prove to be plausible candidates for an expression of psycho-physical dependence. The problem with global supervenience is that it is too restrictive. It allows the possibility that if two possible worlds differ with respect to some minute physical detail (for instance, Saturn's rings contain one more ammonia molecule), they may differ radically with respect to mental properties.[15] Such a relation between mental and physical properties does not suggest what one should expect from psycho-physical dependence.[16] As a dependency relation weak supervenience fares no better. The problem with weak supervenience is that it lacks the modal force required to generate dependence between the related properties:

> Determination or dependence is naturally thought of as carrying a certain modal force: if being a good man is dependent on, or is determined by, certain traits of character, then having these traits must insure or guarantee being a good man (or lacking certain of these traits must insure that one not be a good man). The connection between these traits and being a good man must be more than a de facto coincidence that varies from world to world.[17]

Without a necessary connection between the supervenient properties and the supervenience base, then, it seems there is little reason to think of the supervening properties as depending on the base properties. Kim claims that Davidson has said he accepts something like weak

---

[14] Similar conclusions follow from concerns about meaning and one's linguistic community. See Tyler Burge, "Individualism and the Mental," in Midwest Studies in Philosophy IV: Studies in Metaphysics, ed. P. French et al. (Minneapolis: University of Minnesota Press, 1979).

[15] Jaegwon Kim, "'Strong' and 'Global' Supervenience Revisited," in Kim (1993) p. 85.

[16] The obvious maneuver of specifying which physical properties in a world are relevant to the distribution of mental properties is blocked by the very fact that the relation is globally defined. Such a move would require an alternative formulation of supervenience, such as local supervenience or something intermediate between local and global supervenience (as is suggested by externalist concerns).

[17] Kim, "Concepts of Supervenience," p. 60.

supervenience,[18] in which case it appears he is correct to suggest that Davidson's characterization of supervenience cannot be regarded as a kind of dependence.[19]

Does the third alternative (strong supervenience) express a relation of dependence? Since strong supervenience ensures more than a de facto coincidence between the related properties (given that the relation holds across possible worlds), one might think that strong supervenience does generate dependence.[20] Kim denies this, however. His reason for this is that dependence appears to be an asymmetric relation and as far as Kim is concerned strong supervenience is "neither symmetric nor asymmetric."[21] Given this, strong supervenience does not appear to be the proper kind of relation to capture what we should intuitively expect from psycho-physical dependence.

> For when we look at the relationship specified in the definition between a strongly supervenient property and its base property, all that we have is that the base property entails the supervenient property. This alone does not warrant us to say that the supervening property is dependent on, or determined by, the base, or that an object has the supervening property in virtue of having the base property. These latter relations hint at an asymmetric relation. We have learned from work on causation and causal modal logic the hard lesson that the idea of causal dependence or determination is not so easily or directly obtained from straightforward modal notions alone; the same in all likelihood is true of the idea of supervenient determination and dependence.[22]

The concern is, then, that the mere fact that the mental supervenes on the physical, even in all possible worlds, is not enough to ground the dependence of the mental on the physical.

---

[18] In his "Replies to Essays X-XII," in Essays on Davidson: Actions and Events, ed. Bruce Vermazen and Merrill B. Hintikka (Oxford: Clarendon Press, 1985) p. 242.

[19] James Klagge has pointed out that it might be a mistake to think of Davidson's characterization of supervenience as equivalent to Kim's weak supervenience, given Davidson's "latent" anti-realism about the mental as it follows from his holism and views on interpretation. See Klagge, "Davidson's Troubles With Supervenience," Synthese 85 (1990) pp. 339-352. I agree with Klagge's suggestion and will make use of it later.

[20] Of course, if it turns out that strong supervenience is equivalent to global supervenience, as some claim, then it is not a good candidate for an expression of dependence either. This issue, however, appears to remain undecided.

[21] Kim, "Concepts of Supervenience," p. 67.

[22] Ibid.

Therefore, such dependence does not, as one might hope, follow from the definition of strong supervenience alone.

Another way of expressing this problem is to say that strong supervenience is consistent with what William Seager calls "correlative" as opposed to "constitutive" supervenience.[23] Correlative supervenience asserts a mere correlation between two families of properties. Thus, correlative supervenience is consistent not only with epiphenomenalism (understood as an account of the mind-body relation), but also with views such as parallelism. Constitutive supervenience, on the other hand, involves the claim that the physical base properties in some sense constitute the supervenient properties. Such constitution, however it is to be understood (I will look at some suggestions later), serves nicely as the ground for a dependency relation and rules out forms of ontological dualism. Since it appears that an assertion of strong supervenience alone does not allow us to distinguish between the constitutive and correlative varieties, it cannot be regarded as an expression of dependence without bringing in some further considerations. This is why some authors, including Kim, have said that supervenience is not a solution to the mind-body problem, but instead expresses the very problem itself. Thus, I take Kim's concern that strong supervenience is non-symmetric to express the worry that it might be a relation of correlative as opposed to constitutive supervenience.[24] This is why Kim, in his later paper, renames his definitions of supervenience weak and strong "covariance." For the formal definitions of supervenience, since they are consistent with correlative supervenience, assert a mere property covariation.

---

[23] William Seager, Metaphysics of Consciousness (London: Routledge, 1991) p. 177.

[24] Similar concerns have led some authors to formulate even stronger modal variants of the relation by adding another necessity operator. For example see Terence Horgan, "From Supervenience to Superdupervenience: Meeting the Demands of a Material World," Mind 102 (1993) pp. 554-586; Thomas Grimes, "Supervenience, Determination, and Dependency," Philosophical Studies 62 (1991) pp. 81-92.

One might try to circumvent this problem by arguing that although asymmetry doesn't follow directly from the definition of strong supervenience itself, it will follow if supervenience builds in concerns for multiple realization (which it seems it should anyhow). Thus, it might appear as though one could characterize strong psycho-physical supervenience as an asymmetric relation by making the property covariation asymmetric: While supervenient A-properties strongly covary with B-properties (base properties), such that whatever is B-indiscernible is A-indiscernible, it might be the case that B-properties don't strongly covary with A-properties; that is, there may be objects that are A-indiscernible but B-discernible. This type of asymmetric covariance seems quite plausible and is required by the principle of multiple realization. If the sensation of pain, for example, can be physically realized in different ways by different organisms, then although all such organisms can share the same mental state of feeling pain, they will not share the same physical state when in pain. Hence we have a form of asymmetric supervenience: A supervenes on B, but B does not supervene on A.

While it is true that the property covariation is asymmetric this is still not enough to ensure the dependence of A-properties on B-properties. Although we can be reasonably sure that B-properties don't depend on A-properties, it doesn't follow from the covariance alone that A-properties depend on B-properties. However, one might argue that the dependence of A on B is the best explanation for the asymmetric covariation. In this case, though, the captured sense of dependence follows not from the property covariation alone, but is postulated as the explanation for the noticed covariation. The problem with this approach is that there might be some other explanation for the asymmetric property covariation, so there is no reason to think that the dependence of A on B is the best explanation available. Kim himself points out this possibility:

> What this argument neglects, rather glaringly, is the possibility that an explanation of the covariance from A to B may be formulated in terms of a third set of properties. It seems clearly possible for there to be three sets of properties A, B,

> and C, such that A and B depend on C, A covaries with B but B does not covary with A, and A does not depend on B. Something like this could happen if, although both A and B covary with C, B makes finer discriminations than A, so that indiscernibility in regard to B-properties entails indiscernibility with respect to A, but not conversely.[25]

For example, Kim says that intelligence (A) strongly covaries with manual dexterity (B), but that manual dexterity does not strongly covary with intelligence. We should not take intelligence to depend upon manual dexterity, however, because it is more likely that both of these characteristics depend on genetic and developmental factors (C).[26]

There are a couple of things that are troublesome about Kim's reply. First, Kim chooses a poor example to illustrate his point. This is because it is unclear precisely how talk about manual dexterity makes "finer discriminations" than talk about intelligence, in which case it is unlikely we would think that intelligence depends on manual dexterity in the first place. More problematic still, the covariance between intelligence and manual dexterity can hardly be regarded as strong covariance, which is the kind of covariance at issue. In fact, it is unlikely that these properties even weakly covary, given that one need not look far for counterexamples to their covariance in this world (Stephen Hawking, for example, is very intelligent, though not manually dexterous). Of course we can't make too much of the fact that Kim has chosen a poor example. His general point remains a possibility; it is unclear that one can rule out the idea of a third property that underlies and explains strong psycho-physical covariance a priori.[27] The idea that there might be a third property responsible for the asymmetric covariation once again takes us back to the possibilities of parallelism and epiphenomenalism, so asymmetric property covariation

---

[25] Kim., "Supervenience as a Philosophical Concept," p. 146.

[26] Ibid.

[27] It bears observing that Seager also raises the possibility of a common cause as a source of correlative supervenience. See Seager, Metaphysics of Consciousness, p.182.

alone will not get us psycho-physical dependence. What if we could somehow rule out the possibility of the third property? Would dependence follow then?

As it turns out, even if we successfully build in concerns for multiple realization and rule out the hypothesis of the third property we will still not have an asymmetric relation when dealing with strong supervenience. The reason for this lies in the way Kim conceives of the supervenience base. According to Kim the base is to be understood as a (possibly) infinite disjunction of properties (including the multiple physical bases for the mental properties of all organisms). By characterizing the base in this way he rules out the possibility of asymmetric covariance.

Here is Kim's argument for the non-asymmetric character of strong supervenience: Strong supervenience says that

> whenever a supervening property P is instantiated by an object, there is a subvenient property Q such that the instantiating object has it and the following conditional holds: necessarily if anything has Q, then it has P. So the picture we have is that for a supervenient property P, there is a set of properties, Q_1, Q_2, . . . in the subvenient set such that each Q_i is necessarily sufficient for P. Assume that this list contains all the subvenient properties each of which is sufficient for P. Consider then their disjunction: Q_1 or Q_2 or . . . (or UQ_i, for short). This disjunction may be infinite; however, it is a well-defined disjunction, as well defined as the union of infinitely many sets. It is easy to see that this disjunction is necessarily coextensive with P.
>
> First, it is clear enough that UQ_i entails P, since any disjunct does. Second, does P entail UQ_i? Suppose not: something then, say b, has P but not UQ_i. According to strong [supervenience], b has some property in the subvenient set, say S, such that necessarily whatever has S also has P. But then S must be one of the Q_i, and since b has S, b must have UQ_i. So P entails UQ_i.[28]

Kim's argument shows why strong supervenience between properties cannot be asymmetric. The relation cannot be asymmetric because the entailment relations between the supervenient properties and the base run in both directions, and since the relation is not asymmetric there can, according to Kim, be no relation of dependence expressed by the property covariation alone. As Kim remarks, "what must be added to covariation to yield

---

[28] Kim, "Supervenience as a Philosophical Concept," pp. 151-152.

dependence is an interesting, and metaphysically deep, question."[29] So it seems my suspicions are correct. The relation of strong supervenience does not yield dependence because without building in other claims about the relation it is consistent with correlative supervenience. However, Kim draws a further lesson from the above argument. He claims that since the supervenient and base properties stand in a relation of mutual entailment they are coextensive, and hence, we have a serviceable bridge law for reduction.[30]

One might think that if the properties of one domain can be reduced to the properties of another all worries about dependence evaporate. True, perhaps, but there are different ways to understand the nature of this "evaporation" depending on one's attitudes toward the connection between reduction and dependence. On the one hand one might think that to show that mental properties are reducible to physical properties expresses a form of dependence because such dependence is already included in the idea of reduction. I suppose the idea here is that if mental properties are identical to physical properties, then they necessarily depend on them, for any property depends on itself for its identity because any property is identical with itself. Something like this seems to be behind William Seager's unexplained claim that identity is the strongest possible form of constitutive supervenience, where constitutive supervenience is meant to capture the idea of dependence.[31] One might also think that Kim accepts something like this idea given his

---

[29] Ibid., p. 148.

[30] Some differ on this interpretation of Kim's view. For example, Ausonio Marras thinks Kim is concerned with a relation between predicates since he interprets the reducibility involved in strong supervenience to be a reduction of one theory to another. However, he also suggests that one can read Kim as describing an ontological reduction. See Ausonio Marras, "Supervenience and Reducibility: an Odd Couple," and "Psychophysical Supervenience and Nonreductive Materialism," Synthese 95 (1993) pp. 275-304. For the view that Kim's reduction is primarily ontological see John Bacon, "Supervenience, Necessary Co-extensions, and Reducibility," Philosophical Studies 49 (1986) pp. 163-176.

[31] Seager, Metaphysics of Consciousness, p. 188.

claim that "to be reduced is to be legitimized."[32] If this is his general attitude toward reduced properties, then the fact that strong supervenience entails the reduction of mental to physical properties does not mean that mental properties are eliminated, in which case he might reasonably claim that mental properties depend on physical properties in the attenuated sense above. Of course this is not the asymmetric form of dependence Kim originally insisted upon, but I see no reason in general why we must be limited to that intuitive form of the relation.

The other attitude, which is more plausible in my opinion, is to say that the question of dependence evaporates because there are no longer distinct items to stand in that relation. That is, if mental properties are to depend on physical properties, then they must necessarily be distinct. Since Kim thinks strong supervenience entails the reducibility of mental properties to physical properties—the reduction is not of one theory to another; Kim says, "We are not here talking about predicates, or linguistic expressions, but properties . . ."[33]—it would seem that mental properties are not distinct from the physical properties on which they supervene, and hence cannot depend on them.[34] In this case, then, the issue of dependence evaporates because it no longer makes sense to say that mental properties depend on physical properties. If this is the case, then we need to look to other forms of non-reductive supervenience if we want to capture the idea of psycho-physical dependence. Let us assume, however, that even if strong supervenience entails reducibility it does express some (albeit attenuated) form of dependence. Can such a reduction be defended?

While Kim's argument is of course logically valid, a point with which one might take issue is his claim that the disjunctive base properties can figure in reductive bridge

---

[32] Kim, "Epiphenomenal and Supervenient Causation," in Kim (1993) p. 95.

[33] Kim, "Supervenience as a Philosophical Concept," .p. 152.

[34] Ausonio Marras expresses such a concern in his "Supervenience and Reducibility: an Odd Couple," Philosophical Quarterly 43 (1993) pp. 215-222.

laws such that one side of the bridge law mentions a mental property and the other side mentions a disjunction of physical properties. I myself find the role of disjunctions in laws to be highly suspicious. Jerry Fodor, in "Special Sciences"[35] argues quite convincingly against the use of disjunctions in laws. Although Fodor's argument is directed at the model of intertheoretic reduction proposed by Putnam and Oppenheim,[36] the main force of his objection works equally well against Kim's proposed property reduction.

While Fodor's point is basically intuitive, it is an intuition that is difficult to resist. If the law of a special science ($S_1x \Rightarrow S_2x$) connects two properties or predicates S_1 and S_2, and these predicates are each coextensive with a respective disjunction of properties or predicates:

$$P_1 \text{ v } P_2 \text{ v } \ldots P_n \text{ and } P^*_1 \text{ v } P^*_2 \text{ v } \ldots P^*_n$$

then the following is a law:

$$P_1 \text{ v } P_2 \text{ v } \ldots P_n \Rightarrow P^*_1 \text{ v } P^*_2 \text{ v } \ldots P^*_n$$

But this is implausible. As Fodor says:

> I think, for example, that it is a law that the irradiation of green plants by sunlight causes carbohydrate synthesis, and I think that it is a law that friction causes heat, but I do not think that it is a law that (either the irradiation of green plants by sunlight or friction) causes (either carbohydrate synthesis or heat).[37]

The objection is thus to the form of the law, regardless of its content. Such statements simply do not appear to be laws.

William Seager arrives at the same conclusion by different means.[38] His concern is that laws ought to be confirmed by their instances but disjunctive laws like Kim's are not. While his argument involves some fairly complex use of probability theory, the basic point

---

[35] Jerry Fodor, "Special Sciences," in The Philosophy of Science, ed. Boyd, Richard, P. Gasper, and J.D. Trout (Cambridge: MIT Press, 1991).

[36] Paul Oppenheim and Hilary Putnam, "Unity of Science as a Working Hypothesis," in Boyd et al. (1991).

[37] Ibid., p. 437.

[38] See also David Owens, "Disjunctive Laws," Analysis 49 (1989) pp. 197-202.

is quite simple. The fact that a certain mental state is confirmed to supervene on a certain neural state for humans tells us nothing about whether the same mental state supervenes on the other physical disjuncts mentioned in the bridge law. But if laws are confirmed by their instances, then this instance should tell us something about the other physical realizations of the mental state at issue. Thus, such "laws" are not, properly speaking, confirmed by their instances and so should not be treated as genuine laws. As Seager puts it, somewhat comically:

> It is presumably a law that sodium burns with a yellow flame. The basic evidence for this is provided by cases of ignited sodium burning yellow. Does this evidence confirm the 'law' that when you light sodium either it burns yellow or gremlins dance a jig on the CN tower?[39]

If we follow Seager and Fodor in their evaluation of the disjunctive laws invoked by Kim to reduce mental properties to the physical properties on which they strongly supervene, then there is little reason to identify mental and physical properties on the basis of the supervenience relation alone. Without this identification strong supervenience ceases to express even the attenuated sense of metaphysical dependence. For we end up with the possibility that strong supervenience is a form of correlative rather than constitutive supervenience, which re-opens the possibilities of epiphenomenalism and parallelism.

The upshot of the discussion to this point is that Kim has leveled some very persuasive arguments against the idea that supervenience, in any of its recognized forms, expresses psycho-physical dependence. Weak supervenience is lacking the appropriate modal force, global supervenience is too restrictive, and strong supervenience lacks the necessary feature of asymmetry. While strong supervenience is certainly better off than its companions, the only sort of dependence it comes close to expressing is the attenuated sense that follows from the reduction of mental to physical properties. Since there are

---

[39] Seager, Metaphysics of Consciousness, pp. 129-130.

compelling reasons to doubt that such a reduction follows from strong supervenience, however, even this weak sense of dependence is not captured by strong supervenience. In the end, then, it appears as though none of the formulations of supervenience we have considered express dependence.

As I mentioned above, some authors have thought that we can avoid the shortcomings of the existing formulations of supervenience by modifying the type and scope of the necessity operators in the definitions of the relation. However, it should be clear from the above analysis of the problems with Kim's original formulations of supervenience that the result will merely strengthen the property covariation rather than explain it. Thus, even modally reinforced definitions of supervenience can only capture the idea of property covariation and so cannot rule out parallelism and epiphenomenalism. Therefore, in what follows I propose a different approach to this problem.

Throughout his discussions of whether or not supervenience expresses dependence Kim has characterized the relation as one that holds between families of properties.[40] This was the case even when he discussed weak supervenience, which he took to be equivalent to Davidson's formulation of the relation. My discussion in Chapter Two of how many criticisms of Davidson's theory misfire because they fail to take seriously his reluctance to endorse talk about properties should alert us to the possibility that the same difficulty arises for Kim's criticism of Davidson's account of supervenience. Perhaps if we conceive of the relation as one that holds between predicates rather than between properties, as Davidson would demand, the difficulties Kim has identified do not arise.

---

[40] In fact, Kim states a number of times that it doesn't matter for the purpose of his discussions whether one conceives of the relation as one between properties or predicates.

Since Davidson's anomalous monism (which includes the supervenience thesis[41]) is a brand of identity theory, one might wonder, as with Kim's, why the issue of dependence comes up at all. The reason it does is that, unlike Kim, Davidson does not identify mental and physical properties. The only identity asserted by Davidson holds between mental and physical events. Therefore, the question remains whether on Davidson's brand of supervenience mental properties are dependent on physical properties, and if so what sort of dependence this expresses. The answers to these questions become clear when we fully appreciate an important difference between Kim's formulation of supervenience and Davidson's.

Kim's thesis is a metaphysical one about the relations between the properties that (as we saw in Chapter Two) constitute events. If we construe Davidson's talk about properties as talk about predicates, as I suggested earlier, then it appears that Davidson's thesis is quite different. Far from a metaphysical thesis, Davidson's is a semantic thesis about our use of language. This is corroborated by the following passage where Davidson tries to clarify what his supervenience thesis should be taken to express:

> the notion of supervenience, as I have used it, is best thought of as a relation between a predicate and a set of predicates in a language: a predicate p is supervenient on a set of predicates S if for every pair of objects such that p is true of one and not of the other there is a predicate of S that is true of one and not of the other.[42]

If we take Davidson seriously here and resist the usual temptation to ignore the difference between properties and predicates, this opens up a way of regarding supervenience as a relation of dependence. However, the dependence gleaned is not metaphysical dependence, it is instead a kind of semantic dependence.

---

[41] Davidson has the supervenience thesis play an even greater role in anomalous monism than it used to. For example, see "Thinking Causes," in Mental Causation, ed. John Heil and Alfred Mele (Oxford: Clarendon Press, 1995).

[42] Donald Davidson, "Replies to Essays," in Essays on Davidson: Actions and Events, ed. B. Vermazen and M. Hintikka (Oxford: Clarendon Press, 1985) p. 242.

The sort of semantic dependence at issue is not the straight-forward variety of showing that certain predicates from different areas of discourse are definitionally equivalent, and hence, can be reduced one to the other in the way some have thought that moral predicates are analytically definable in terms of naturalistic predicates. Davidson explicitly rules out this possibility when he says in his description of supervenience that it "does not entail reducibility through law or definition . . . (emphasis added)."[43] What other sort of dependence might there be?

I think the answer to this question lies in Davidson's interpretationalism. According to Davidson, to have beliefs and desires is to have them ascribed by an interpreter: "If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything."[44] For Davidson the ascription of beliefs is part and parcel of the process of radical interpretation. One is not possible without the other. The starting point for such an interpretation must be the physical behaviour of the agent in question. Thus, what we can say about the physical state of the agent is primary in the enterprise of interpretation, as are the physical conditions of the environment when an utterance is made (for instance, the passing rabbit when the agent utters "gavagai"). Similarly, we require behavioural (hence physical) evidence to ascribe a change in belief state to the speaker. This is implicit in the second half of Davidson's definition of supervenience in "Mental Events": "an object cannot alter in some mental respect without altering in some physical respect. . . ."[45] If one follows this idea through to its logical conclusion I think it is apparent that Davidson's treatment of the mental must

---

[43] Donald Davidson, "Mental Events," p. 214.

[44] Davidson, "Radical Interpretation," in Davidson, Inquiries into Truth and Interpretation (Oxford: Clarendon Press, 1984) p. 137.

[45] Donald Davidson, "Mental Events," in Davidson, Essays on Actions and Events (Oxford: Oxford University Press, 1980) p. 214.

include the idea that the mental is dependent on the physical in a broad sense, namely, that what we can say about someone's beliefs is determined by what physical predicates can be ascribed to that person. Furthermore, if one takes Davidson's externalism seriously and thinks of it as analogous to Putnam's, so that elements of the physical world are partly constitutive of one's mental states, then this idea is hard to resist. Thus, I think there is good reason to say that for Davidson mental predicates are determined by, and dependent on, physical predicates. This may seem to be a strange suggestion in light of the so-called circle of intentionality that has come to be associated with Davidson's holism about the mental, but when one considers the ground for mental ascription I think there is good reason to believe that the autonomy of mental descriptions is frequently overstated in Davidson scholarship.

The worry that a number of authors have that Davidson's version of supervenience entails that two people who are identical in every respect except for one seemingly irrelevant physical detail must have different beliefs (e.g., one person has one eyelash that is longer than his or her counterpart's)[46] begins to seem unfair to Davidson if we read him in the way I have suggested. For this worry proceeds from the false assumption that there is a particular set or subset of physical predicates (for instance, ones describing the brain rather than eyelashes) responsible for any given mental state. If we take Davidson's holism seriously this just isn't so. Hence, it appears that these kinds of concerns about Davidson's theory have missed their mark and have ignored Davidson's externalism.

One might wonder if the same difficulties that plagued Kim's formulations of supervenience might resurface for Davidson's. In particular, since Davidson's thesis seems to be modally weak, why don't mental and physical predicates vary in their covariation

---

[46] For example see Simon Evnine, Donald Davidson (Stanford: Stanford University Press, 1991) pp. 69-70.

across possible worlds to the point where their covariation in this world is a mere "de facto coincidence"? Such a worry seems to me to be misguided and to miss the point of my suggested interpretation of Davidson's thesis. We have seen that mental predicates depend on physical predicates in a straight-forward sense: To have beliefs and desires is to have them ascribed by an interpreter on the basis of physical facts about the speaker. The relation between mental and physical predicates, then, is not one of mere covariation in this world. It is only if the opposite were true that one would be tempted to modalize the relation in order to generate dependence. Thus, given the way that Davidsonian supervenience has been conceptualized, it should be evident that the fact that Davidson's version of supervenience is modally weak does not create a problem with regarding it as a relation of dependence.

The interpretation of Davidson's version of supervenience I have proposed certainly gives us a sense of psycho-physical dependence, but many will find it dissatisfying. I suspect the reason for this is that most would prefer the kind of metaphysical dependence between mental and physical properties Kim tries to develop. Kim's thesis is more exotic than Davidson's and the potential for reduction is, to many, very attractive from a physicalist point of view. However, as we saw the prospect of property reduction is quite implausible. Furthermore, it proceeds from the confusing premise that there are such things as mental properties, which, despite being allegedly reducible to physical properties, have an unspecified and somewhat mysterious status of their own. If we proceed from Davidson's assumption that events are mental only when described using mental predicates, and predicates are simply components of a language which obey the rules of language use and interpretation, then much (though not all) of the mystery of the mental is removed. Thus, I suggest that rather than try to squeeze metaphysical dependence out of psycho-physical supervenience by formulating ever stronger modal variants of the relation,

we follow Davidson's lead and think of the relation as one expressing semantic dependence.

2. Supervenience and Explanatory Completeness

Now that I have shown that supervenience can be seen to express a form of psycho-physical dependence the groundwork for my defense of physicalism is in place. All that remains to be done is to show precisely how this relation can be used to block the epiphenomenalist challenges and thereby salvage the explanatory completeness of physicalism. I model my argument on a different but related discussion by Jaegwon Kim. In "Mechanism, Purpose, and Explanatory Exclusion,"[47] Kim examines the question of whether or not an explanation of an agent's action in terms of mechanism (neuro-physiological processes, for instance) precludes an explanation of the same token of behaviour in terms of the agent's reasons for acting. Although Kim accepts the principle of explanatory exclusion, which claims that there cannot be more than one complete and independent explanation for any event, Kim argues for a negative answer to this question. In his view psychological explanation supervenes on mechanistic explanation and thereby depends on the latter. Since the principle of explanatory exclusion holds only for complete and independent explanations, the dependence of psychological explanation on mechanistic explanation shows that the mechanistic explanation does not exclude a psychological account of the same token of behaviour. Although my intention in this section is somewhat different from Kim's, I want to suggest that the same argument can be used to different ends: it shows that reason explanations and facts about qualia are not autonomous as they

---

[47] Jaegwon Kim, "Mechanism, Purpose, and Explanatory Exclusion," in Kim (1993).

first appeared, in which case the consequences the autonomy of these phenomena would have for physicalism do not follow. As we shall see, however, the details of the argument will have to be altered slightly in light of the conclusions drawn in the previous section. While Kim's understanding of supervenience is insufficient to reach the desired conclusion because it does not capture an appropriate sense of dependence, the alternative Davidsonian version of supervenience will.

Kim's attempt to justify the legitimacy of what I have called "reason explanations" is motivated in large part by an assumption, common among physicalists, that there is something inherently wrong with having multiple explanations for any phenomenon, especially when there is a commitment to the idea that there is a basic physical explanation for any event. Andrew Melnyk describes this view as follows:

> Suppose you are a physicalist, so that you believe (at least) that every event, without exception, just is some fundamental-physical event. It seems to follow that every event, without exception, has a fundamental-physical explanation. But if every event, without exception, has a fundamental-physical explanation, then every event, without exception, has an explanation. What, therefore, is the point of the explanations apparently supplied by the special sciences, by the sciences distinct from fundamental physics? They seem, indeed, quite needless, since they explain nothing that is not already explained. But if, like explanations citing phlogiston, they are explanatorily dispensable, surely we should dispense with them, just as we have dispensed with the phlogiston-citing explanations. After all, if everything is physical, and everything physical has a physical explanation, then nothing would be left unexplained if we were to junk all special scientific explanations.[48]

In order to go about reconciling the claim that there can be multiple explanations for any particular phenomenon with the above assumption that physicalism is explanatorily complete, one must first appreciate in more detail what the resistance to the idea of multiple explanations consists in. The view that multiple explanations are unacceptable has its roots in what has come to be called "the principle of explanatory exclusion," first argued for by

---

[48] Andrew Melnyk, "Testament of a Recovering Eliminativist," Philosophy of Science (Proceedings) 63 (1996) p. S185.

Norman Malcolm[49] and taken up more recently and defined by Jaegwon Kim as follows: "No event can be given more than one complete and independent explanation."[50]

In its early form, Malcolm argued that there could not be two distinct explanations for one action, such as a man's climbing a ladder. The alternative explanations he considered are (1) a mechanistic explanation in terms of physiology and (2) what I have called a "reason explanation" in terms of the agent's mental states such as his reasons for acting. Malcolm concluded that to admit the truth of a mechanistic account excludes the truth of the alternative reason explanation. His reasons for adopting this principle stem primarily from worries about causal overdetermination. If both explanations identify different sufficient conditions for the behaviour in question, then we have a case of causal overdetermination because the explanations identify two distinct causes of the ladder climbing. It is important to recognize, however, that causal overdetermination as such is not necessarily a problematic notion. In fact, the stock example of someone being both fatally shot and poisoned is an example of causal overdetermination we can imagine and understand quite well. The thing to notice about this example is that even though the poison and the bullet fired from the gun are each sufficient for the death of our unfortunate victim, an explanation of the victim's death can easily incorporate both causal elements as partial causes of the victim's death. In this sense, then, separate explanations in terms of the gunshot and the poisoning are not complete even though each of the partial causes is sufficient on its own for the mentioned effect. The fact that these two stories (in terms of the gunshot and the poison) can be combined in an intelligible causal explanation of the victim's death is crucial to making sense of the overdetermination. As Malcolm saw it, this

---

[49] Norman Malcolm, "The Conceivability of Mechanism," Philosophical Review 77 (1968) pp. 45-72. For a direct reply to Malcolm's argument see Alvin Goldman, "The Compatibility of Mechanism and Purpose," Philosophical Review 78 (1969) pp. 468-482.

[50] Jaegwon Kim, "Mechanism, Purpose, and Explanatory Exclusion," in Kim (1993) p. 239.

sort of unified account cannot be told in the case of an intentional action. One cannot work the reason for the agent's action into the physiological account of behaviour in the way one can presumably combine the causal elements of the poison and the fired bullet into one explanation. To combine the two accounts of the ladder climbing into one explanation would require that we fit the reason somewhere into the same causal story as the physiological account of the ladder climbing, and to do so gives us an unacceptable schism in our explanation.[51] The picture we end up with is one of purely physical forces at work in the body being causally influenced by some mental event which is not part of the same causal system as the physiological events that make up the bulk of the mechanistic explanation. This, of course, bears an uncomfortable resemblance to the sort of account Descartes offered of action-explanation, whereby mental events mysteriously intervene in the chain of physical events in the body via the pineal gland.

Since I have argued that reason explanations are not a species of causal explanation but that the reasons identified in such explanations nevertheless have causal powers, it is perhaps unclear why the principle of explanatory exclusion is directly relevant to my concerns in this section. After all, if reason explanations represent a distinct explanatory category from causal explanations, then it seems fairly clear that there is no need to worry about competing causes in cases like the one Malcolm identified, in which case the exclusion principle is neatly averted. In fact, Kim himself suggests as much in a footnote:

> [O]ne way in which one might try to eliminate the incompatibility is to interpret rationalizing explanation as a fundamentally noncausal mode of understanding actions. I believe that this is an approach well worth exploring: a rationalizing

---

[51] Malcolm assumes that no psycho-physical identity theory is true, so one cannot appeal to an identity between the reason for acting and one of the physical states of the body that figures in the causal account of the ladder climbing to get around this problem.

> explanation is to be viewed as a normative assessment of an action in the context of the agent's relevant intentional states.[52]

Thus, since the worries about explanatory exclusion are, at least for Malcolm, derived primarily from concerns about causal overdetermination, it would appear as though the principle can hold only for alternative causal explanations. Since reason explanations belong to a different category of explanation, and the principle of explanatory exclusion appears to hold only for competing explanations of the same species, it would seem as though the principle of exclusion is not an issue of concern to this dissertation.

While there is good reason to believe that Kim thinks reason explanations represent a distinct explanatory category from causal explanations, and that the former fits with the account I offered of psychological explanation in the context of my discussion of Davidson, I have doubts about how seriously Kim takes the suggested implications this has for the debate about the principle of explanatory exclusion. A number of remarks suggest that Kim thinks the principle holds even for alternative explanations of different species.

The first reason for thinking this is that each time Kim defines the principle (including the definition quoted above) he does not say "causal explanations"; instead he speaks generally of "explanations."[53] While it is possible that it is implied that he has in mind causal explanations rather than explanations of this and other species, I think his general remarks about the epistemology of explanation makes this very implausible. According to Kim, explanations are provided in order to improve our epistemic standing or to solve certain epistemic "predicaments" and in his view "too many explanations will put us right back into a similar epistemic predicament" unless we can show how the alternative

---

[52] Kim, "Mechanism, Purpose, and Explanatory Exclusion," in Kim (1993) p. 240. Kim elaborates this view of rationalizing explanation in "Self-Understanding and Rationalizing Explanations," Philosophia Naturalis 21 (1984) pp. 309-320.

[53] Kim, "Mechanism, Purpose, and Explanatory Exclusion," in Kim (1993) pp. 239, 250, 257.

explanations are related to one another.[54] Motivating this belief is a general view of explanation as simplification and unification. If we are to explain something well we should do so without multiplying entities or explanatory premises beyond necessity, or if we do have multiple accounts we should strive to explain how the accounts are related according to an underlying unifying principle. I am not here concerned to defend this view of explanation. The important point for the present discussion is that it is a view Kim endorses. Given this, multiplying our explanations of any phenomenon without accounting for the relations between them leads us into confusion, for in this case we have no complete and unified story to tell about the phenomenon in question; at best we have several fragmentary or partial explanations, and since the explanations remain unconnected they appear to be in conflict with one another; we are left wondering how they can all be correct, or indeed whether some of them might be false. Kim summarizes this view as follows:

> If simplicity and unity of theory is our aim when we seek explanations, multiple explanations of a single phenomenon are self-defeating—unless, that is, we are able to determine that their explanatory premises are related to one another in appropriate ways.[55]

I take this to express a fundamental attitude toward explanation which holds regardless of the species of explanation involved, for Kim says that unity and simplicity "are general considerations not restricted to causal explanations"[56] Thus, it seems clear that Kim's principle of explanatory exclusion should hold even for explanations of different species. In this case, though, the worry motivating the principle is not the same as Malcolm's (i.e., overdetermination or competing causes). Instead, the rationale is that alternative explanations lead us away from a satisfactory epistemic situation by failing to meet general standards of unity and simplicity. Whether or not these are good reasons to reject multiple

---

[54] Ibid., p. 254.
[55] Ibid.
[56] Ibid., p. 255.

explanations is unimportant. The significant thing for our purposes here is to note that Kim is committed to the view that alternative explanations of the same phenomenon, even if the explanations belong to different explanatory categories, are to be avoided.

While Kim accepts the principle of explanatory exclusion he denies that it holds in the case Malcolm identifies. The reason for this is that Kim believes the two explanations are not independent explanations as Malcolm assumes. Since he thinks that the explanation of behaviour in terms of mental states is not independent of the explanation in terms of mechanism, both explanations can be admitted without violating the principle of explanatory exclusion.

Kim identifies two possible ways of showing this. Either (1) the explanations of behaviour in terms of mental states are reducible to physical explanations, or (2) they depend on physical explanations. In Kim's view both possibilities are available if one accepts the view that the mental supervenes on the physical. So even if Malcolm has reasons for rejecting the type-identity theory, supervenience, despite having the appearance of a relatively weak version of physicalism, provides a connection between our two explanations that is strong enough to avoid the principle of explanatory exclusion. With the first option, if reason explanations are reducible to physical explanations, then where we thought we had two explanations we in fact have only one, and so there can be no worry about competing explanations. With the second option, if reason explanations are dependent on physical explanations, then although we stop short of saying that we in fact have only one explanation we nevertheless avoid the principle of exclusion because the principle applies only to independent explanations. By claiming that reason explanations depend on physical explanations the physical is given the ontological and explanatory priority required by any form of physicalism, in which case reason explanations are no longer autonomous, and hence, no longer threaten physicalism. For the dependence of the reason explanation on a more detailed causal explanation in physical terms suggests that the

explanation in physical terms is a "deeper and more inclusive story of how the behaviour came about."[57]

Although Kim briefly mentions that the first alternative is a genuine possibility, he does not explicitly argue for it in the context of his discussion of Malcolm's argument. This is probably just as well since, as we saw in the first section, it is doubtful that any brand of supervenience (even strong supervenience) entails the reducibility of supervenient to base properties. For it is unlikely that one can formulate the bridge laws necessary to reduce the mental to the physical in light of the disjunctions required by the principle of multiple realizability. Since the property reduction fails there is little reason to expect that the two explanations (in terms of the respective sets of properties) collapse into one explanation.

Instead of taking the reductionist route, Kim claims that the relation of supervenience is itself sufficient to get around the problem of explanatory exclusion. Kim believes this for two related reasons. First, he assimilates mental causation to macro causation (causation involving ordinary, medium sized physical bodies), which, he claims, is a form of supervenient causation.[58] That is to say, Kim thinks that mental causation is no different from instances of ordinary causal relations such as fire causing smoke, billiard balls moving one another, and so on. Like these ordinary causal relations, mental causation supervenes on more complex microphysical causal relations. Second, by virtue of the fact that mental causation <u>is</u> supervenient causation, and so supervenes on microphysical causation, he thinks that mental causation depends on microphysical causation. This means that the explanations of behaviour in terms of intentional states depend on basic microphysical causal explanations, and such dependence is sufficient to avoid the principle of explanatory exclusion.

---

[57] Ibid., p. 241.

[58] See Jaegwon Kim, "Epiphenomenal and Supervenient Causation," in Kim (1993).

It is not necessary here to go too deeply into the details of Kim's account of supervenient causation. The basic point is clear enough. For him all macro causation is supervenient causation. Furthermore, it is epiphenomenal supervenient causation. It is "epiphenomenal" in the sense that all macro causal relations are apparent rather than real since the actual causal connections between events hold between the basic physical properties that constitute them and on which the macro properties supervene. Thus, Kim's account of epiphenomenal causation is epiphenomenal in the sense that he and many other critics think Davidson's account of mental causation is epiphenomenal. The real causal work is being done by underlying physical properties, and since mental properties supervene on such properties this creates the illusion that there is a causal relation between mental properties and actions, but this is only an illusion. Nevertheless this legitimizes all macro causal explanations, including intentional explanations of behaviour since such explanations are grounded in more basic physical explanations, and hence, might be regarded as abbreviated (but non-equivalent) accounts of the more complete explanations that would be provided by neurophysiology or physics.

While this seems a plausible way of rescuing the explanatory completeness of physicalism, the conclusions drawn in the previous section appear to block off this sort of approach. For we have seen that unless one can show that strong supervenience entails the reducibility of the mental to the physical it does not express a relation of dependence when the supervenience relation is thought of as one that holds between properties, and even then this captures only an attenuated sense of dependence. Recall that given the denial of reduction, strong supervenience is consistent with correlative supervenience, and so there is little reason to suppose that strong supervenience expresses a relation of dependence. Since the possibility of reduction is undermined, and Kim has ruled out the possibility that his other forms of supervenience express dependence, Kim has no resources available to him to support his claim that supervenient explanations depend on micro physical

explanations. Given this, it appears that Kim cannot avoid the principle of explanatory exclusion in the way he suggests.

As I pointed out earlier in this chapter, Davidson's version of supervenience is much more promising than Kim's as far as providing a relation of dependence between the mental and the physical goes. Perhaps if we turn to Davidson we might (in a manner similar to that Kim attempted) be able to show using his own characterization of supervenience that reason explanations do not violate the principle of explanatory exclusion. If successful, then we will have shown that the apparent autonomy of psychological explanation is not as threatening to physicalism as it appears to be.

One might think that since, for Davidson, explanation is always explanation under a description, the worries I have about the autonomy of psychological explanation and the principle of explanatory exclusion are completely misguided. In "Thinking Causes," Davidson briefly discusses Kim's worries about explanatory exclusion and claims as much. He says,

> The idea [of explanatory exclusion] is that if physics does provide . . . 'full, sufficient' explanations, there is no room for mental explanations unless these can be (fully, strictly?) reduced to physical explanations. What can this strange principle mean? If we consider an event that is a 'full, sufficient' cause of another event, it must, as Mill pointed out long ago, include everything in the universe preceding the effect that has a causal bearing on it . . .; and even then, if we take 'sufficient' seriously, we must assume perfect determinism. How can the existence of such an event 'exclude' other causes? It can't, since by definition it includes everything that could be a cause.[59]

Furthermore, Davidson suggests that even if we could provide such complete physical explanations of events this does not necessarily preclude other explanations (e.g., reason explanations), for explanation, unlike causation, is "interest-sensitive"; the way an explanation functions depends on what our interests are and how an event is described.

---

[59] Donald Davidson, "Thinking Causes," in Mental Causation, ed. John Heil and Alfred Mele (Oxford: Clarendon Press, 1993) pp. 15-16.

Davidson's point is that given the intensional character of explanation, in contrast to the extensional character of causation, there is no reason to suppose that there is any conflict between alternative explanations. The cause of an event can be picked out under infinitely many descriptions, and the salient features of these descriptions may fit into antecedently held information in different ways, yielding logically alternative explanations. It is a mistake, Davidson urges, to confuse causation with causal explanation and to expect that different descriptions of a cause thereby identify distinct causes:

> It is only if we confuse causal relations, which hold only between particulars, with causal explanations, which, so far as they are 'sufficient' must deal with laws, and so with types of events, that we would be tempted to accept the principle of 'causal-explanatory exclusion'.[60]

So in Davidson's view, Kim's worries about multiple explanations are misguided because they stem from his confusing causal relations, which hold between concrete particular events, and causal explanations, which involve the formulation of law-like statements connecting descriptions of events, and thus require descriptions of events that fall under appropriate kinds, be they psychological kinds, biological kinds, or whatever.

While Davidson has a good point, I think he is somewhat unfair to Kim on this matter. First, I have suggested that Kim's motivations for accepting the principle of explanatory exclusion do not necessarily stem from worries about causal overdetermination, in which case Davidson's insistence that we not confuse causation with causal explanation is beside the point. Second, Davidson places undue emphasis on the completeness of explanations. True, the principle of explanatory exclusion holds for alternative complete explanations, but Kim leaves it quite open-ended exactly what this sense of "completeness" should involve, in which case Davidson might very well have attributed too strong a sense of "completeness" to Kim's principle. More importantly, such

---

[60] Ibid., 16.

explanations must also be independent, and Davidson says little about the question of dependence which figures so prominently in Kim's discussion. The one remark Davidson does make about the dependence of intentional explanations on physical explanations appears in a footnote just after the passage quoted above. He says,

> I have . . . [neglected the condition of independence] because dependence means entirely different things in the cases of events and of explanation. Events 'depend' on one another causally, and the failure of psycho-physical laws has no bearing on the question of whether mental and physical events are causally related. Explanation, on the other hand, is an intentional [sic] concept; in explanation, dependence is geared to the ways in which things are described. There is no reason why logically independent explanations cannot be given of the same event . . . .[61]

While I have no quarrel with Davidson's remarks here, they do demonstrate a failure on his part to recognize the seriousness of the problem his view faces. My discussion in Chapter Two showed that intentional explanation cannot, given the anomalism of the mental, be construed as a species of causal explanation, and without a means of grounding reason explanations in more basic causal explanations that can be expressed using physical predicates, the explanatory completeness of physicalism is undermined because reason explanations thereby begin to seem autonomous. If one can show that such explanations depend on more basic physical explanations in the way Kim suggests, then this will remove the threat of explanatory autonomy and thereby rescue the explanatory completeness of physicalism. The general question Kim raises about relations of dependence between explanations (rather than events), then, is not entirely confused as Davidson suggests. In fact, if I am correct, it can be of tremendous help to Davidson's position.

Davidson does take up this point in an indirect fashion. Davidson's central aim in "Thinking Causes" is to show how many of his critics have gone wrong in thinking that anomalous monism leads to epiphenomenalism. At one point Davidson presses his thesis

---

[61] Ibid., footnote 9, p. 16.

of supervenience into use to account for how the mental properties of an event can be seen to "make a difference" to causal relations.

> For supervenience as I have defined it does, as we have seen, imply that if two events differ in their psychological properties, they differ in their physical properties (which we assume to be causally efficacious). If supervenience holds, psychological properties make a difference to the causal relations of an event, for they matter to the physical properties, and the physical properties matter to causal relations.[62]

While I think this way of putting things is unfortunate and is bound to cause further confusion regarding Davidson's account of mental causation, I think we can draw some useful lessons from this claim if we remember what Davidson's attitudes are toward mental properties. What is required to show that the mental properties of an event that figure in a reason explanation are dependent on the event's physical properties is the reverse of the argument just quoted. Not only do the mental properties of an event "make a difference" to the event's causal relations, but an event's physical properties (which figure in causal explanations) make a difference to that event's mental properties, particularly those properties that figure in reason explanations. This is included in the second (and frequently ignored) part of Davidson's definition of supervenience in "Mental Events": "an object cannot alter in some mental respect without altering in some physical respect." As I suggested toward the end of the previous section, this should be understood as expressing a form of semantic dependence since the relation is one between mental and physical predicates. The idea here is that since Davidson is an antirealist about mental properties, one ought to understand his talk about "properties" as talk about predicates, as linguistic items that can be assigned truly to agents in sentences about their behaviour. Thus, mental predicates are ascribed to agents in accord with general principles of interpretation. Part of what must be considered in the interpretation of someone's behaviour (even one's own) is

---

[62] Ibid., p. 14.

the environment and physical state of the individual. In fact, such considerations are primary in the task of mental ascription. Thus we are justified in ascribing a change in belief state to someone only when we have behavioural evidence to do so, and this, of course, requires physical changes for us to work with. Without such changes there is no evidence for us to appeal to in order to justify ascribing a mental difference to the agent.

By establishing a relation of dependence in this way between mental and physical predicates it appears that the pressure on anomalous monism identified in Chapter Two is greatly alleviated. For given this we can appreciate how intentional explanations depend on the causal explanations possible in the language of physics. Intentional explanations depend on what physical predicates are true of some event or relation, and since physical predicates figure in causal explanations (of varying degrees of precision), intentional explanations depend on causal explanations. This relation of dependence shows that reason explanations are not independent explanations and hence do not violate the principle of explanatory exclusion. Reason explanations are therefore grounded in an appropriate way in causal explanations so that it remains true that all psychological facts and explanations depend on and are determined by physical facts as physicalism requires. To distinguish, as I did in Chapter Two, between reason explanations and causal explanations in no way undermines Davidson's brand of physicalism. The distinction still holds, but it does not follow from this distinction that reason explanations are autonomous.

The same reasoning removes the problems with qualia identified in Chapter Three. Recall that the worry was that although qualia are physical phenomena their qualitative characteristics cannot be captured in physical terms. Because there appear to be facts that escape capture in physical terms it seems as though physicalism is explanatorily incomplete. But if one adopts the view that the mental is supervenient on the physical, this holds no less for qualia than it does for propositional states. In this case, then, although qualia cannot be captured in physical terms, facts about qualia are nevertheless dependent

on and determined by physical facts. This shows that, as with reason explanations, facts about qualia are not autonomous in a way that is threatening to physicalism.[63]

Supervenience might also be employed to further support the claim in Chapter Three that qualia are, despite their elusiveness to physical theory, physical properties. An analogy with other supervenient properties shows how. Consider the wetness of water. We can reasonably say that the property of wetness supervenes on certain molecular-chemical properties. While the wetness of water is quite evident at the macro level of analysis it is unobservable at the molecular-chemical level. Nevertheless, it seems quite implausible to insist that the wetness of water is thereby a non-physical property. Similarly, we might suggest that phenomenal properties are properties that are unobservable at the level of neuroscience but are observable at the macro level in the sense I explained earlier: that one needs to be in the requisite brain state to have access to them. As with the wetness of water, the fact that certain properties of brain states are inaccessible at the level of scientific investigation need not entail that such properties are non-physical.

In light of these observations it seems as though there is little reason to suppose that the epiphenomenalist objections to the forms of physicalism we have considered undermine the explanatory thesis of physicalism. We can conclude, then, that the epiphenomenalist objections fail completely. They do not succeed in showing the falsity of the explanatory completeness of physicalism.

---

[63] Frank Jackson himself might now accept a view like this. See Jackson, "Mental Causation," Mind 105 (1996) pp. 377-413.

Chapter 5

Conclusions and Further Problems

The examination of the epiphenomenalist objections to physicalism has shown that those objections pose little threat to physicalism. In Chapters Two and Three I described the objections in detail and showed that the strongest possible conclusion that can be drawn from them is that physicalism is incomplete at the explanatory level. In Chapter Four I advanced the second stage of my argument in a discussion about the concept of supervenience and showed that this concept, if understood in a particular way, undermines even the conclusions drawn in the preceding chapters. Thus, the pattern of argument has been as follows:

1. Physicalism (the ontological claim plus the claim to explanatory completeness) is true.
2. There are facts and explanations that do not depend on physical explanations.
3. Therefore, physicalism is not explanatorily complete.
4. Hence, (1) is false.
5. But if we accept the Davidsonian-style thesis of supervenience, then (2) is false.
6. Therefore, (3) and (4) do not follow.

The discussion of supervenience reveals that the argument against physicalism has a false premise. The initial appeal of (2) stems, I think, from an ambiguity about the explanatory inadequacies of the forms of physicalism considered. While it is true that there are facts that cannot be captured in physical terms and that there are explanations that are distinct from

the causal explanations ordinarily associated with the physical sciences, this does not necessarily mean that these facts and explanations do not depend on physical facts and explanations. If physicalism can be supplemented with (or characterized in terms of) a claim about psycho-physical supervenience, then there is little reason to think that premise (2) follows from the identified explanatory inadequacies. In the end, then, it is clear that the traditional epiphenomenalist arguments are of little threat to physicalism, though they are nevertheless worth considering in order to clarify issues of central concern.

The significance of these conclusions is substantial. The epiphenomenalist objections I discussed have long been regarded as significant obstacles to physicalism. To undermine the force of these objections is therefore to emancipate a number of different versions of physicalism and consequently to open the way to more adequate articulations of this view. *However, my discussion in this thesis falls short of providing an adequate* characterization of physicalism. The reason for this is that it was not my intention to articulate and defend any particular version of this view; my aim was primarily a defensive one on behalf of physicalism as a general thesis which might be made more precise in a number of ways. However, it is worth considering in a preliminary and rather sketchy manner what a plausible form of physicalism might look like in light of the conclusions drawn in the previous chapters. This brief discussion should point the way to further research and potential problems that will have to be considered in formulating a fully developed theory of mind.

While both functionalism and Davidson's anomalous monism were shown to be less problematic theories than most have thought, the relation of psycho-physical supervenience was pivotal in completely undermining the epiphenomenalist objections. Thus, I suggest that supervenience must be a part of any adequate physicalist theory. When Davidson made use of supervenience it was as a supplement to anomalous monism. Token-identity is, as he says "consistent with the view that mental characteristics are in some sense

dependent, or supervenient, on physical characteristics."[1] Since the thesis of supervenience was used by Davidson as a supplement to anomalous monism (i.e., it is not entailed by the premises of anomalous monism, but is merely consistent with them), it is clear that this relation can be used by other physicalist theories as well, so long as it is consistent with them also. Thus, it remains an open possibility that functionalism, for example, could incorporate a form of weak supervenience like Davidson's in which case it emerges as a strong contender for an adequate articulation of physicalism.

Of course the best and most obvious candidate for a plausible version of physicalism is Davidson's own theory. This is because we have seen that the thesis of supervenience must be understood in a very specific sense if it is to be construed as a relation of dependence: it must be understood as a relation between predicates rather than properties. Such an account of supervenience requires the anti-realist/interpretationalist attitude toward mental states I attributed to Davidson in Chapter Two. Since it is not clear that functionalists are anti-realists about the mental in this sense, some work would have to be done to determine whether or not such a view of the mental is consistent with the main claims of functionalism. But what about anomalous monism itself? Is such a view consistent enough to represent a plausible theory of the mind?

There are several pressures on Davidson's theory that would need to be addressed in order for it to qualify as a promising articulation of physicalism. The first of these stems from Davidson's anti-realist stance toward the mental. Since it follows from Davidson's view that there are no fixed and determinate facts about someone's beliefs (i.e., that there are no facts of the matter about mental content), Davidson's theory is much closer to eliminative materialism than most have thought. For if agents have mental states only as

---

[1] Donald Davidson, "Mental Events," in Davidson, Essays on Actions and Events (Oxford: Oxford University Press, 1980.) p. 214.

ascribed (by themselves and others), and such ascriptions are indeterminate, then what is to stop us from thinking that the use of the intentional idiom is, as Quine once suggested, "essentially dramatic"?[2] It is a small step from Davidson's interpretationalism to the claim that people don't really have mental states and that such ascriptions are simply a matter of practical convenience. Like Dennett's Intentional Stance,[3] one can claim that mental states are ascribed to agents in order to predict and explain behaviour. This means of predicting and explaining behaviour is far more efficient, though perhaps much less accurate, than that provided by neurophysiology or physics. But the temptation associated with such a view is to say that people don't really have beliefs, desires, and other mental states. While appeals can be made to practical indispensability (since too much information would be required to predict and to explain behaviour on a day to day basis if we were to limit ourselves to the language of neurology), there is little to prevent the philosophical conclusion that the mental is simply a convenient fiction.

I think it is a simple matter to arrive at this conclusion if one regards Davidson as an anti-realist about the mental as I have suggested. If this conclusion seems unsavory, however, the alternative is even worse. To regard Davidson as a realist about mental properties is, as many critics have argued, to open himself to the charge of epiphenomenalism, in which case physicalism is undermined.[4] Thus, I think it is preferable either to come to terms with the apparent consequences of anti-realism about mental properties or else find ways to resist the slide toward eliminativism. I suggested one means of resisting this conclusion in Chapter Four. One might argue that the move to eliminate the mental proceeds from the assumption that intentional and neurological explanations of

---

[2] W. V. O. Quine, Word and Object (Cambridge: MIT Press, 1960) p. 219.

[3] See Daniel Dennett, The Intentional Stance. Cambridge: MIT Press, 1989.

[4] That is, if one can avoid the idea that events have their properties essentially as I pointed out in Chapter Two.

behaviour are in competition, and that this is a false assumption. My discussion of supervenience and the principle of explanatory exclusion supports this idea, but more work would have to be done to show that eliminativism can be permanently held at bay. This is work that goes well beyond the reaches of the present study, but is certainly worth pursuing.

A second, related pressure on a Davidsonian account concerns the apparent discrepancy between the propositional attitudes and the qualitative states of consciousness. While many might be comfortable with the anti-realist/interpretationalist view of propositional attitudes, it is not clear that one can view qualia in the same manner, especially if one suggests, as I have, that there is an element of perceptual experience that can be known only by acquaintance. Since qualia are, in a significant (though qualified sense) subjective, it is difficult to assimilate our understanding of such items to the kind of treatment Davidson proposes for the propositional attitudes. In particular, it does not appear to be a matter of interpretation that the ripe tomato I perceive involves acquaintance with a quale of a particular sort. Since supervenience was employed to save the explanatory thesis of physicalism where qualia are concerned, but supervenience must be construed in a Davidsonian sense, it is not clear that it can actually work in the case of qualia as I suggested it might.

I see two possible strategies for dealing with this difficulty. The first possibility is to adopt what William Seager calls "differential supervenience."[5] The central idea here is that different kinds of mental states supervene differently on physical states. So while intentional states might weakly supervene on physical states, qualia might involve a different form of supervenience. I do not pretend that my discussion of Kim's account of

---

[5] See William Seager, Metaphysics of Consciousness (London: Routledge, 1991) p. 114.

supervenience and dependence incorporated every conceivable formulation of supervenience. There might be many others.[6] However, in order for an alternative form of supervenience to successfully undermine the challenge to the explanatory completeness of physicalism any such alternative must capture the sense of dependence required to avoid correlative supervenience. Davidson's version of supervenience does this admirably. It is, given Kim's problems, unclear how versions of supervenience that are not articulated in terms of predicates could achieve this.

The second alternative, although more complicated, might represent a more plausible solution. Rather than try to develop differing accounts of supervenience to deal with the apparent differences between the propositional attitudes and qualia, one might attempt to weaken the differences between these kinds of mental states. If qualia could be shown to be sufficiently like the propositional attitudes, then the apparent differences between them would shrink and there would be less difficulty with applying the Davidsonian version of supervenience uniformly to both. The beginning of such an account was identified in Chapter Three. In my discussion of qualia inversion I suggested, following Simpson, that one might regard the affective component of sensation as partly constitutive of qualitative content. For instance, pain might be seen to be constituted in part by the emotion of self-pity. If, as I suggested, this analysis of sensation could be generalized such that an affective component were identified for each kind of sensation, then it would be possible to describe qualitative states of consciousness in terms of affective content. Since affective content can be expressed in propositional terms, qualia turn out to be more like the propositional attitudes than most have thought. Of course this represents only a rough outline of a potential account of qualia. Filling in the details to give

---

[6] One example of which is suggested by Terence Horgan in his "Supervenient Qualia," Philosophical Review XCVI No. 4 (1987) pp. 491-520.

a complete account would be a difficult and involved task. Nevertheless, if decreasing the differences between qualia and the propositional attitudes is possible, then the problems with applying Davidson's model of supervenience to qualia evaporate and we are left with a uniform account of the relation between mental and physical events and predicates.

We are still a long way from a comprehensive and adequate theory of mind. There is much hard work to do and many difficult problems to solve. Nevertheless, some of the more tenacious obstacles to a physicalist theory of mind have been removed in this dissertation. This is perhaps the most one can hope for in philosophy: to open rather than to close possibilities.

Bibliography

Ackerman, Diane. A Natural History of the Senses. New York: Vintage Books, 1990.

Antony, Louise. "Anomalous Monism and the Problem of Explanatory Force," Philosophical Review XCVIII (1989) pp. 153-187.

Aquinas, St. Thomas. Summa Theologica. Trans. Anton C. Pegis. New York: Random House, 1945.

Armstrong, David. The Nature of Mind and Other Essays. St. Lucia: University of Queensland Press, 1980.

Bacon, John. "Supervenience, Necessary Co-extensions, and Reducibility," Philosophical Studies 49 (1986) pp. 163-176.

Bilodeau, Renée. "L'Inertie du Mental," Dialogue XXXII (1993) pp. 507-524.

Block, Ned. "Troubles with Functionalism," in Readings in Philosophy of Psychology Vol. 1. Edited by Ned Block. Cambridge: Harvard University Press, 1980, pp. 268-305.

--------. "Are Absent Qualia Impossible?" Philosophical Review LXXXIX (1980) pp. 257-274.

Block, Ned and Fodor, J. "What Psychological States are Not," in Block (1980).

Boyd, Richard, P. Gasper, and J.D. Trout (eds.). The Philosophy of Science. Cambridge: MIT Press, 1991.

Burge, Tyler. "Individualism and the Mental," in Midwest Studies in Philosophy IV: Studies in Metaphysics. Edited by P. French et al. Minneapolis: University of Minnesota Press, 1979.

Campbell, Neil. "Aquinas' Reasons for the Aesthetic Irrelevance of Tastes and Smells," British Journal of Aesthetics 36 (April, 1996) pp. 166-176.

--------. "The Standard Objection to Anomalous Monism," Australasian Journal of Philosophy 73 (September, 1997).

--------. "Putnam on the Token-Identity Theory," Philosophia 27 (forthcoming).

Cartwright, Nancy. "The Reality of Causes in a World of Instrumental Laws," in Philosophy of Science Association 1980 Vol. 2. Edited by P. Asquith and R. Giere. East Lansing, MI: Philosophy of Science Association, 1981. Reprinted in Boyd et al.(1990).

Churchland, Paul. "The Logical Character of Action Explanations," Philosophical Review 79 (1970) pp. 214-36.

--------. "Eliminative Materialism and Propositional Attitudes," Journal of Philosophy 78 (1981) pp. 67-90.

--------. Matter and Consciousness. Cambridge: MIT Press, 1984.

--------. "Reduction, Qualia, and the Direct Introspection of Brain States," Journal of Philosophy 82 (1985) pp. 8-28.

--------. "Folk Psychology and the Explanation of Behavior," in The Future of Folk Psychology. Edited by John Greenwood. Cambridge: Cambridge University Press, 1991.

Churchland, Paul and Churchland, Patricia. "Functionalism, Qualia and Intentionality," in Mind, Brain and Function: Essays in the Philosophy of Mind. Edited by J. Biro a nd R. Shahan. Norman, OK: University of Oklahoma Press, 1982.

Conee, Earl. "Physicalism and Phenomenal Qualities," Philosophical Quarterly 35 (1985) pp. 296-302.

Cornman, J.W. Materialism and Sensations. New Haven: Yale University Press, 1971.

Crane, Tim and D. H. Mellor. "There is no Question of Physicalism," Mind 99 (1990) pp. 185-206.

Damasio, A., Lima, A., and Damasio, H. "Nervous Function After Right Hemispherectomy," Neurol. 25 (1975) pp. 89-93.

Davidson, Donald. "Actions, Reasons, and Causes," Journal of Philosophy 60 (1963). Reprinted in Davidson (1980).

--------. "Mental Events," in Experience and Theory. Edited by Lawrence Foster and J. W. Swanson. Mass.: University of Massachusetts Press and Duckworth, 1970. Reprinted in Davidson (1980).

--------. "On the Very Idea of a Conceptual Scheme," Proceedings and Addresses of the American Philosophical Association 47 (1974). Reprinted in Davidson (1984).

--------.Essays on Actions and Events. Oxford: Oxford University Press, 1980.

--------. Inquiries into Truth and Interpretation. Oxford: Clarendon Press, 1984.

--------. "Replies to Essays X-XII" in Essays on Davidson: Actions and Events. Edited by Bruce Vermazen and Merrill B. Hintikka. Oxford: Clarendon Press, 1985.

--------. "Thinking Causes," in Mental Causation. Edited by John Heil and Alfred Mele. (Oxford: Clarendon Press, 1993).

Dennett, Daniel. Brainstorms. Cambridge: MIT Press, 1981.

--------. The Intentional Stance. Cambridge: MIT Press, 1989.

--------. Consciousness Explained. Boston: Little Brown, 1991.

Dretske, Fred. "What Good is Consciousness," Canadian Journal of Philosophy 27 (March, 1997) pp. 1-15.

Dumouchel, Paul. "Ce que l'on Peut Apprendre Sur les Chauves-Souris à l'Aide d'une Télé Couleur," Dialogue XXXII (1993) pp. 493-505.

Evnine, Simon. Donald Davidson. Stanford: Stanford University Press, 1991.

Flanagan, Owen J., Jr. The Science of the Mind. Cambridge: MIT Press, 1984.

--------. Consciousness Reconsidered. Cambridge: MIT Press, 1992.

Fernald, Anne. "Intonation and Communicative Intent in Mothers' Speech," Child Development 60 (December, 1989) pp. 1497-1510.

--------. "Prosody and Focus in Speech to Infants and Adults," Developmental Psychology 27 (March, 1991) pp. 209-221.

Fernandez, Ephrem and Dennis C Turk. "Sensory and Affective Components of Pain: Separation and Synthesis," Psychological Bulletin 112, No. 2 (1992) pp. 205-217.

Fodor, Jerry. "Special Sciences, or The Disunity of Science as a Working Hypothesis," Synthese 28 (1974) pp. 77-115. Reprinted under the title "Special Sciences," in Boyd et al. (1990).

Goldman, Alvin. "The Compatibility of Mechanism and Purpose," Philosophical Review 78 (1969) pp. 468-482.

Greenwood, John (ed.) The Future of Folk Psychology: Intentionality and Cognitive Science. Cambridge: Cambridge University Press, 1991.

Grimes, Thomas. "The Myth of Supervenience," Pacific Philosophical Quarterly 69 (1988) pp. 125-160.

--------. "Supervenience, Determination, and Dependency," Philosophical Studies 62 (1991) pp. 81-92.

Hacking, Ian. "Weapons Research and the Form of Scientific Knowledge," in Nuclear Weapons, Deterrence and Disarmament. Edited by D. Copp. Calgary: University of Calgary Press, 1986.

Hare, R. M. The Language of Morals. London: Oxford University Press, 1952.

Harvey, Jean. "Systematic Transposition of Colours," Australasian Journal of Philosophy 57 (1979) pp. 211-219.

Heil, J. and A. Mele (eds.). Mental Causation. Oxford: Clarendon Press, 1995.

Hempel, Carl. "Reasons and Covering Laws in Historical Explanation," in Philosophy and History: A Symposium. Edited by Sidney Hook. New York: New York University Press, 1963.

--------. "Laws and Their Role in Scientific Explanation," in Hempel, Philosophy of Natural Science Chapter 5. Englewood Cliffs, New Jersey: Prentice Hall, 1966. Reprinted in Boyd et al. (1991).

Hess, Peter. "Actions, Reasons, and Humean Causes," Analysis 40 (1981) pp. 77-81.

--------. Thought and Experience. Toronto: University of Toronto Press, 1988.

Hoffman, Paul. "Cartesian Passions and Cartesian Dualism," Pacific Philosophical Quarterly 17 (1990) pp. 310-333.

Honderich, Ted. "The Argument for Anomalous Monism," Analysis 42 (1982) pp. 59-64.

--------. "Anomalous Monism: Reply to Smith," Analysis 43 (1983) pp. 147-149.

--------. "Smith and the Champion of Mauve," Analysis 44 (1984) pp. 86-89.

Horgan, Terence. "Supervenience and Micro-physics," Pacific Philosophical Quarterly 63 (1982) pp. 27-43.

--------. "Jackson on Physical Information and Qualia," Philosophical Quarterly 34 (1984) pp. 147-152.

--------. "Supervenient Qualia," Philosophical Review XCVI (1987) pp. 491-520.

--------. "From Supervenience to Superdupervenience: Meeting the Demands of a Material World," Mind 102 (1993) pp. 554-586.

Huxley, T. H. "On the Hypothesis That Animals Are Automata," in Huxley, Methods and Results: Essays. London: Macmillan co., 1901.

International Association for the Study of Pain, "Classification of Chronic Pain: Descriptions of Chronic Pain Syndromes and Definition of Pain Terms," Pain Suppl. 3 (1986) S1-S226.

Jackson, Frank. "Epiphenomenal Qualia," Philosophical Quarterly 32 (1982) pp. 127-136.

--------. "What Mary Didn't Know," Journal of Philosophy 83 (1986) pp. 291-295.

--------. "Mental Causation," Mind 105 (1996) pp. 377-413.

--------. "The Primary Quality View of Colour," in Philosophical Perspectives Vol. 10. Edited by James E. Tomberlin. Oxford: Blackwell, 1996.

Johnsen, Bredo. "Dennett on Qualia and Consciousness: a Critique," Canadian Journal of Philosophy 27 (March, 1997) pp. 47-82.

Kandel, Eric R. and James H. Schwartz (eds.). Principles of Neural Science (2nd edition). New York: Elsevier Science Publishing Co., Inc., 1985.

Kazez, Jean R. "Can Counterfactuals Save Mental Causation?" Australasian Journal of Philosophy 73 (1995) pp. 71-90.

Kim, Jaegwon. "Events as Property Exemplifications," in Action Theory: Proceedings of the Winnipeg Conference on Human Action.Edited by M. Brand and D. Walton. Dordrecht: Reidel, 1976.

--------. "Psychophysical Supervenience," Philosophical Studies 41 (1982) pp. 51-70.

--------. "Self-Understanding and Rationalizing Explanations," Philosophia Naturalis 21 (1984) pp. 309-320.

--------. "Concepts of Supervenience," Philosophy and Phenomenological Research 45 (1984) pp. 153-176. Reprinted in Kim (1993).

--------. "Epiphenomenal and Supervenient Causation," Midwest Studies in Philosophy 9 (1984) pp. 257-270. Reprinted in Kim (1993).

--------. "Psychophysical Laws," in Actions and Events: Perspectives on the Philosophy of Donald Davidson. Edited by Ernest LePore and B. McLaughlin. New York: Basil Blackwell, 1985.

--------. "'Strong' and 'Global' Supervenience Revisited," Philosophy and Phenomenological Research 48 (1987) pp. 315-326. Reprinted in Kim (1993).

--------. "Mechanism, Purpose, and Explanatory Exclusion," in Philosophical Perspectives 3, Philosophy of Mind and Action Theory. Edited by James E. Tomberlin. Atascadero, Cal.: Ridgeview Publishing Co., 1989. Reprinted in Kim (1993).

--------. "Supervenience as a Philosophical Concept," Metaphilosophy 21 (1990) pp. 1-27. Reprinted in Kim (1993).

--------. Supervenience and Mind, Selected Philosophical Essays. Cambridge: Cambridge University Press, 1993.

--------. "Can Supervenience Save Anomalous Monism?" in Heil and Mele (1995).

Kincaid, Harold. "Supervenience and Explanation," Synthese 77 (1988) pp. 251-281.

Kirk, R. "From Physical Explicability to Full-Blooded Materialism," Philosophical Quarterly 29 (1979) pp. 229-237.

Kivy, Peter. The Corded Shell: Reflections on Musical Expression. Princeton: Princeton University Press, 1980.

Klagge, James. "Davidson's Troubles With Supervenience," Synthese 85 (1990) pp. 339-352.

Leibniz, Gottfried. "Monadology," in Leibniz: Philosophical Writings. Edited by G. H. R. Parkinson. Translated by Mary Morris and G. H. R. Parkinson. London: J.M. Dent and Sons Ltd., 1973.

Lennon, Kathleen. "Reduction, Causality, and Normativity," in Reduction, Explanation and Realism. Edited by D. Charles and K. Lennon. Oxford: Clarendon Press, 1992.

LePore, Ernest and B. McLaughlin (eds.). Actions and Events: Perspectives on the Philosophy of Donald Davidson. New York: Basil Blackwell, 1985.

Levenhal, H. and D. Everhart. "Emotion, Pain, and Physical Illness," in Emotions in Personality and Psychopathology. Edited by C. E. Izard. New York: Plenum Press, 1979.

Lewis, David. "Mad Pain and Martian Pain," in Block (1980).

--------. "Attitudes De Dicto and De Se," Philosophical Review 88 (1979) pp. 513-543.

--------. "Postscript to 'Mad Pain and Martian Pain'," in Lewis, Philosophical Papers Vol. 1. Oxford: Oxford University Press, 1983.

Lewis, M. M. Infant Speech: A Study of the Beginnings of Language. London, Routledge & Kegan Paul, 1936/1951.

Locke, John. An Essay Concerning Human Understanding. Edited by Peter H. Nidditch. New York: Oxford University Press, 1975.

Lowe, E. J. "The Causal Autonomy of the Mental," Mind 102 (1993) pp. 629-644.

Ludwig, Kirk A. "Causal Relevance and Thought Content," Philosophical Quarterly 44 (1994) pp. 334-353.

Lycan, William. "Form, Function, And Feel," Journal of Philosophy (1981) pp. 24-50.

Madell, Geoffrey. "Physicalism and the Content of Thought," Inquiry 32 (1989) pp. 107-121.

Malcolm, Norman. "The Conceivability of Mechanism," Philosophical Review 77 (1968) pp. 45-72.

Marras, Ausonio. "Materialism, Functionalism and Supervenient Qualia," Dialogue XXXII (1993) pp. 475-492.

--------. "Supervenience and Reducibility: An Odd Couple," Philosophical Quarterly 43 (1993) pp. 215-222.

--------. "Psychophysical Supervenience and Nonreductive Materialism," Synthese 95 (1993) pp. 275-304.

--------. "Nonreductive Materialism and Mental Causation," Canadian Journal of Philosophy 24 (1994) pp. 465-494.

--------. "The Debate on Mental Causation: Davidson and His Critics," Dialogue XXXVI (1997) pp. 177-195.

McDowell, John. "Functionalism and Anomalous Monism," in Ernest LePore and B. McLaughlin (1985).

McGinn, Colin. The Problem of Consciousness. Cambridge: Basil Blackwell, 1991.

--------. "Conceptual Causation: Some Elementary Reflections," Mind 100 (1991) pp. 573-586.

Mellor, D. H. "Materialism and Phenomenal Properties," Aristotelian Society Supp. 47 (1973) pp. 107-119.

Melnyk, Andrew. "Physicalism: From Supervenience to Elimination," Philosophy and Phenomenological Research LI (1991) pp. 573-587.

--------. "Testament of a Recovering Eliminativist," Philosophy of Science (Proceedings) 63 (1996) pp. S185-S193.

Mesnet, E. "De l'Automatisme de la Mémoire et du Souvenir, dans le Somnambulisme Pathologique," L'Union Médicale Juillet 21 et 23, 1874.

Moore, G. E. Philosophical Studies. London: Oxford University Press, 1922.

Nagel, Thomas. "What Is It Like to Be a Bat?" Philosophical Review 83 (1974) pp. 435-450.

Nemirow, Laurence. "Physicalism and the Cognitive Role of Acquaintance," in Mind and Cognition: A Reader. Edited by W. Lycan. Cambridge: Cambridge University Press, 1990.

Oppenheim, P. and H. Putnam. "Unity of Science as a Working Hypothesis," in Minnesota Studies in the Philosophy of Science Volume II. Edited by H. Feigl, M. Scriven, and G. Maxwell. Minneapolis, MN: University of Minnesota Press, 1958. Reprinted in Boyd et al. (1991).

Owens, Joseph. "Disjunctive Laws," Analysis 49 (1989) pp. 197-202.

--------. "Content, Causation, and Psychophysical Supervenience," Philosophy of Science 60 (1993) pp. 242-261.

Place, U. T. "Is Consciousness a Brain Process?" British Journal of Psychology 47 (1956) pp. 44-50.

Putnam, Hilary. "Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind," Journal of Philosophy XCI (1994) pp. 445-517.

--------. Representation and Reality. Cambridge: MIT Press, 1988.

Quine, W. V. O. Word and Object. Cambridge: MIT Press, 1960.

---------. Pursuit of Truth. Cambridge: Harvard University Press, 1990.

Ramsey, W., S. Stich, and J. Garon. "Connectionism, Eliminativism, and the Future of Folk Psychology," in Greenwood (1991).

Raymont, Paul. "Tye's Criticism of the Knowledge Argument," Dialogue XXXIV (1995) pp. 713-726.

Robinson, Don. "On Crane and Mellor's Argument Against Physicalism," Mind 100 (1991) pp. 135-136.

Rorty, Richard. Philosophy and the Mirror of Nature. Princeton: Princeton University Press, 1979.

Saidel, Eric. "Content and Causal Powers," Philosophy of Science 61 (1994) pp. 658-665.

Seager, William. "The Anomalousness of the Mental," Southern Journal of Philosophy 19 (1981) pp. 389-401.

--------. "Weak Supervenience and Materialism," Philosophy and Phenomenological Research 48 (1988) pp. 697-710.

--------. Metaphysics of Consciousness. London: Routledge, 1991.

--------. "Critical Notice of Fred Dretske, Naturalizing the Mind," Canadian Journal of Philosophy 27 (March, 1997) pp. 83-109.

Searle, John. "Minds, Brains and Programs," Behavioural and Brain Sciences 3 (1980) pp. 417-458.

-------- The Rediscovery of the Mind. Cambridge: MIT Press, 1992.

Shalkowski, Scott. "Supervenience and Causal Necessity," Synthese 90 (1992) pp. 55-87.

Sherrington, C. S. "Cutaneous Sensations," in Textbook of Physiology. Edited by E. A. Schafer. London: Pentland, 1900.

Shoemaker, Sydney. "Functionalism and Qualia," in Block (1980).

--------. "Absent Qualia Are Impossible-A Reply To Block," Philosophical Review XC (1981) pp. 581-599.

--------. "The Inverted Spectrum," Journal of Philosophy LXXIX (1982) pp. 357-381.

Simpson, Evan. "Sensation Deconstructed," in Entities and Individuation: Studies in Ontology and Language in Honour of Neil Wilson. Edited by Donald Stewart. New York: Edwin Mellin Press, 1989.

Skillen, Anthony. "Mind and Matter: a Problem That Refuses Dissolution," Mind 93 (1984) pp. 514-526.

Smart, J. J. C. "Sensations and Brain Processes," Philosophical Review 68 (1959) pp. 141-156.

Smith, Peter. "Hess on reasons and Causes," Analysis 41 (1981) pp. 206-209.

--------. "Bad News for Anomalous Monism?" Analysis 42 (1982) pp. 220-224.

--------. "Anomalous Monism and Epiphenomenalism: A Reply to Honderich," Analysis 44 (1984) pp. 83-86.

Stalnaker, Robert. "Varieties of Supervenience," in Philosophical Perspectives Vol. 10. Edited by James E. Tomberlin. Oxford: Blackwell, 1996.

Stoutland, Frederick. "Oblique Causation and Reasons for Action," Synthese 43 (1980) pp. 351-367.

--------. "Davidson on Intentional Behavior," in Lepore and McLaughlin (1985).

Tye, Michael. "The Subjective Qualities of Experience," Mind 95 (1986) pp. 1-17.

Thornton, Mark. Folk Psychology: An Introduction. Toronto: Canadian Philosophical Monographs, 1989.

Thorp, John. "Astérix at les Qualia: la Dernière Poche de Résistance," Dialogue XXXII (1993) pp. 462-473.

Velleman, David J. "What Happens When Someone Acts?" Mind 101 (1992) pp. 461-481.

Vermazen, B. and Hintikka, M. (eds.). Essays on Davidson: Actions and Events (Oxford: Clarendon Press, 1985).

Weiskrantz, L. Blindsight: A Case Study and Implications. Oxford: Clarendon Press, 1986.

White, Peter A. "Ideas About Causation in Philosophy and Psychology," Psychological Bulletin 108 (1990) pp. 3-18.

Wittgenstein, Ludwig. Philosophical Investigations. Translated by G. E. M. Anscombe (3rd ed.). Oxford: Basil Blackwell, 1968.

Yablo, Stephen. "Mental Causation," Philosophical Review 101 (1992) pp. 245-280.