A DEFENSE OF NON-REDUCTIVE PHYSICALISM

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Matthew Christian Haug

May 2007

A DEFENSE OF NON-REDUCTIVE PHYSICALISM

Matthew Christian Haug, Ph. D.

Cornell University 2007

I develop a novel formulation of, and argument for, non-reductive physicalism – roughly, the view that mental properties are natural properties that are realized by, but not identical to, neural and other low-level physical properties. Non-reductive physicalism has long been the dominant view in the philosophy of mind but has recently been challenged from two main directions. The first type of attack, the causal exclusion problem, points out an apparent inconsistency in non-reductive physicalism. The second type of attack focuses on the multiple realizability of mental properties: questioning either its prevalence or its efficacy in blocking reduction.

In response to the exclusion problem, I first argue that one of the claims used to formulate the problem, the completeness of physics, has two parts and that there is no single domain of physical entities that is the smallest domain of which both parts are true. The conflation of these two parts has made it appear that non-reductive physicalism is inconsistent. I then show how to use the two completeness claims as part of an argument for a form of physicalism that need not be reductive.

In response to the second type of attack, I provide a novel basis for the irreducibility of mental properties. I argue that irreducibility is ultimately grounded in relations between mechanisms, of which multiple realizability is merely one facet. The other facet, multiple determinativity – in which a single physical property realizes several different kinds of high-level properties – is equally effective at blocking property reduction. Thus, even if the doubts about the multiple realizability of mental properties (and its efficacy in blocking reduction) were sound, this would not undermine non-reductive physicalism. Another virtue of this framework is that it

provides an adequate metaphysical basis for some of non-reductive physicalism's explanatory claims – e.g., that high-level explanations are sometimes deeper and theoretically more fecund than low-level physical explanations.

BIOGRAPHICAL SKETCH

Matthew Christian Haug was born in Topeka, Kansas on March 10, 1978, to Melissa Gail Haug and John Alfred Haug, Jr. He attended public schools in the Shawnee Heights school district and graduated from the University of Kansas in May 2000 with a B.A. in philosophy and a B.S. in mathematics. While spending a year "off" taking more philosophy classes and working part-time at the KU Center for Research, he applied for several fellowships and to graduate school in philosophy.

He moved to Ithaca, New York in the autumn of 2001. While not reading and writing about philosophy, he enjoyed hiking, biking, and cross-country skiing in the hills and glens of upstate New York. He earned a M.A. in philosophy in 2004 and will be Assistant Professor of philosophy at the College of William and Mary in the fall of 2007.

For Laurelin

ACKNOWLEDGMENTS

Thanks to the members of the Sage School of Philosophy, past and present, for showing me how to do philosophy, both by their example and through their instruction and conversation. I would like to express my deepest gratitude to the members of my special committee – Dick Boyd, Mike Fara, Sydney Shoemaker, and Nick Sturgeon – for their guidance and for their penetrating comments on drafts of these chapters. Special thanks go to Dick, the chair of my committee, for many illuminating conversations (and many cups of coffee). Thanks also to my fellow graduate students for discussions of matters philosophical and otherwise. I am especially grateful to Andrew Alwood, Emily Esch, Eric Gilbertson, Eric Hiddleston, Paul Kelleher, Anne Nester, Raul Saucedo, and Peter Sutton for helpful conversations about topics that appear in this dissertation (in one form or another).

I am also grateful to Gregory Janssen, Christine Porter, and Helen Steward, for comments on earlier versions of some of this material at conferences (the 2004 meeting of the Creighton Club, the 2005 University of Washington Graduate Student Conference, and the 2005 Oxford Philosophy Graduate Conference, respectively).

This material is based upon work supported under a National Science Foundation Graduate Research Fellowship. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation. Thanks also to the Andrew W. Mellon Foundation, the Graduate School at Cornell University, and the Sage School of Philosophy for fellowship support.

Thanks to the staff at the Granite Mountains Desert Research Center, Mojave National Preserve, California, for an idyllic environment in which to write and think. Finally, thanks to Laurelin Evanhoe for the opportunity to philosophize, botanize, and do ecology in the desert, and for so much more.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Non-reductive physicalism has long been the consensus view about the metaphysics of mind and of the special sciences. As a version of physicalism, it maintains, roughly, that all empirical entities are physical and that all empirical properties supervene on or are realized by physical properties. However, what is distinctive about non-reductive physicalism is the claim that physicalism is compatible with the causal efficacy of irreducible special science properties. For example, non-reductive physicalists claim that mental properties are distinct from microphysical, neurological or biochemical properties while remaining causally efficacious. In general, non-reductive physicalists claim that some perfectly natural properties are not fundamental physical properties, i.e., not properties invoked by a theory of fundamental physical reality.

In the past decade or so, the non-reductive physicalist consensus has been attacked by two related lines of argument: (1) the causal exclusion problem, according to which it is inconsistent to hold that mental properties are causally efficacious and irreducible to physical properties while at the same time maintaining that the physical domain is causally complete and that mental events do not causally overdetermine physical effects, and (2) doubts regarding realization's and multiple realizability's efficacy in arguments for non-reductive physicalism. This second line of argument takes a variety of forms. Some philosophers allege that an adequate account of the realization relation leads to the conclusion that the multiple realizability argument for irreducibility fails – that realization directly provides for a form of reduction (e.g. Kim (1998, 377)). Others have challenged the assumption that all disjunctive properties are unnatural and suggested that a multiple realized property can be reduced to the

disjunction of its possible (non-disjunctive) realizers (e.g. Clapp (2001)). Still others have reexamined what is required for substantive multiple realization and go on to raise doubts about whether mental properties meet these requirements (e.g. Shapiro (2000, 2004), Bechtel and Mundale (1999)).

In this dissertation I respond to these two lines of argument and provide a novel grounding for non-reductive physicalism. In the process I clarify the nature of the debate between reductive and non-reductive physicalists and correct some mistaken assumptions made by both sides of that debate. Consequently, my reformulation of non-reductive physicalism does not fit nicely into the current typology of positions.

I present an account of realization that utilizes the notion of a causal process or mechanism and argue that this mechanistic framework for realization allows us to see why, contra consensus, multiple realizability is not the sole basis for irreducibility (and explicates some puzzling passages from two of the architects of the non-reductive physicalist consensus that seem to suggest this). Multiple realizability is a special case of the true ground of irreducibility: a mismatch between mechanisms involving (families of) realizing properties and realized properties. While "trivial" multiple realizability (which does not involve different mechanisms and in which the differences in the realizers are irrelevant to the instantiation of the realized property) is sufficient to block outmoded, logical empiricist forms of reduction, only substantive multiple realizability is sufficient to block more recent formulations of reduction (such as Jaegwon Kim's "functional" model of reduction, derived from the Ramsey/Lewis approach to defining theoretical terms). Substantive multiple realizability involves a many-one relation between, what I call in Chapter 5, sustaining and integrative mechanisms, respectively. However, even if substantive multiple realizability does not obtain, a one-many relation between sustaining and integrative mechanisms – a

little noticed and unexplored phenomenon that I call "multiple determinativity" – can also block reduction. Finally, I show how this mechanistic framework provides a metaphysical basis for claims non-reductive physicalists make about the explanatory autonomy of the special sciences.

1.1. Some Preliminaries about the Nature of Events and the Distinction between Reductive and Non-Reductive Physicalism

Classic statements of non-reductive physicalism, like Jerry Fodor's (1974), take up the burden of showing how the existence of autonomous special sciences is compatible with the core claims of physicalism. Fodor is concerned with intertheoretic reduction, in which the question is whether natural kind predicates from one theory can be defined in terms of, or are coextensive with, natural kind predicates from another theory (usually fundamental physical theory). The terms of the debate – what is meant by 'reduction' – have shifted since then, and it is not entirely clear how the debate should best be characterized. In this section I argue that the distinction between token-physicalism and type-physicalism regarding events or properties does not adequately characterize the debate. In Chapter 3, I suggest that a disagreement about natural properties does a better job.

One can interpret Fodor's paper as an attempt to salvage "the generality of physics" – the claim that "physics is the basic science" (1974, 97), "that all events which fall under the laws of any science are physical events and hence fall under the laws of physics" – by showing that it does not entail the stronger claim, which he believes to be false, that all sciences reduce to physics. Thus, one might see Fodor's project not as a polemic against all varieties of reductionism but rather as attempting to show what kinds are reduction are required by physicalism and which are plausible given the empirical findings of the special sciences. As he puts it, the goal of scientific reduction is not, as the logical positivists' account of the unity of science has

it, to find a physical natural kind predicate that is co-extensive with each special science natural kind predicate (to reduce all sciences to physics). Rather, it is to "explicate the physical mechanisms whereby events conform to the laws of the special sciences" (1974, 107). The point of Fodor's argument is that the latter does not entail the former, and that the heterogeneity of the physical mechanisms that underlie a given special science law shows that the two goals can come apart. One can achieve the latter goal of explicating physical mechanisms without achieving the former goal of the reduction of natural kinds.

Further, Fodor *accepts* the goal of reduction when it is construed in terms of mechanistic explanation. He is thus best interpreted as arguing for a particular *reformulation* of the logical empiricist doctrine of the unity of science – the claim that physics is general (universally applicable) and is the basic science – a reformulation that does not require bridge laws or principles, property identities, or coextensive natural kind predicates. My project is best interpreted in this way as well. Thus, in the end it is somewhat misleading to emphasize the non-reductive aspect of my formulation of physicalism. There is little point in arguing about what reduction *really* is; many theses fall under that label. But it is important to discuss which of these theses are important, which are plausible, and which can be held independently of one another. Put another way, the label attached to a variety of physicalism – whether it is reductive or non-reductive – is far less important than its content and whether it is well-supported and true. Any form of physicalism must be reductive in one good sense; it must hold that everything empirical is physical in some sense (or as others put it: everything is constituted by, or supervenient on, or realized by the physical).

Calling the position defended in this dissertation "non-reductive" is important only in the context of the current dialectic in the literature. My formulation of

physicalism incorporates a novel argument for some of the key slogans endorsed by prominent non-reductive physicalists: that the special sciences are "autonomous," that some natural kinds are not fundamental physical kinds, that the classifications of the special sciences cross-cut those of fundamental physics. As we shall see, my formulation of non-reductive physicalism is compatible with some claims and methods that have been thought of as reductive by scientists (if not all philosophers) – most notably, it endorses some mechanistic explanations of mental processes. Because of this, it is likely not possible to scientifically investigate mental and other realized properties completely independently of their physical realizers. (To think otherwise is to hang on to the last shreds of the ghost in the machine.) Details of the realization of mental properties (e.g. biological facts) may be relevant to whether or not a mental property is had by a class of organisms. Nevertheless, this does not undermine the autonomy of the special sciences, for similarities at the fundamental physical level do not capture all of the world's causal structure. They miss some objective causal processes/mechanisms. In short, my formulations of non-reductive physicalism can capture a grain of truth found in some reductive forms of physicalism, while showing that this fails to lead to the causal impotence of irreducible mental properties or to the disappearance of (or instrumentalism regarding) the special sciences.

Non-reductive physicalism denies that the weak kind of reduction – the claim that everything is physical – requires a more thoroughgoing, stronger kind of reduction. This stronger form has been formulated in various ways in the past. One traditional formulation is that all scientific terms can be given explicit definitions in fundamental physical terms (see, e.g., (Hellman and Thompson 1975, 551, 556-7, 560-1), (Hempel 1980)). Another is the demand that there be laws or conditionals that bridge the distinct vocabulary of different theories (e.g., (Nagel 1961)). These

formulations are now largely discredited, and those who are part of the recent backlash against non-reductive physicalism are keen to distance themselves from them. Various members of the backlash against non-reductive physicalism endorse "new wave reductionism," (Bickle 1998), "functional reduction" (Kim 2005), or some other form of mind-brain identity theory (Polger 2004).

However, it is not clear what these views have in common, how they differ from older forms of reductionism, and precisely what they find objectionable in non-reductive physicalism. Spelling this out amounts to providing an account of the core difference between reductive and non-reductive physicalism – what I will call "drawing the distinction." A successful proposal for drawing the distinction should meet a couple of criteria. First, it should be *specific*; it should characterize the core of the dispute between these two versions of physicalism and not conflate it with other issues. This goes along with the plausible idea that the debate between reductive and non-reductive physicalism should not hinge on recondite issues about, say, the nature of properties, events, or the causal relation; that is, no matter which metaphysical framework one uses (or no matter which turns out to be correct or most theoretically useful), the core distinction between reductive and non-reductive physicalism should remain unchanged. Second, the proposal should be *informative*; the distinction between reductive and non-reductive physicalism should be substantive – one position should not be a notational variant of the other, and non-reductive physicalism should remain distinct from dualism and strong forms of emergentism (in which configurations of physical entities result in the emergence of novel fundamental forces).

The distinction is often drawn by claiming that both reductive and non-reductive physicalists agree that every token event (or property instance) is identical to a physical event (property instance), but that non-reductive physicalists deny, while

reductive physicalists assert, that every event type (or property) is identical to a physical event type (property) – in short, that non-reductive physicalism can be characterized as endorsing token-identity while denying type-identity (see (Fodor 1974, 100); (Davidson 1980, 1993)).

I claim this is not a successful means of drawing the distinction. Characterizing non-reductive physicalism in this way either fails to be *specific*, by confusing the core dispute between reductive and non-reductive physicalists with more general debates about the ontology of properties and events or fails to be *informative*, by failing to capture the ontological dispute between reductive and non-reductive physicalism, while ensuring that they are both versions of physicalism. The literature in this area is confused and confusing because authors are often talking past one another regarding the nature of events. Theories of events fall into two broad classes: the view that events are property exemplifications and the view that events are concrete particulars (regions of space-time or the contents of those regions). On either account, token-identity conjoined with the denial of type-identity is not an adequate characterization of non-reductive physicalism.

For example, Kim has argued repeatedly that the causal inheritance principle – the claim that the causal powers of a realized property instance are identical with the causal powers of the realizer instance[1] – implies that property instances of multiply realized properties are identical to the instances of the properties that realize them on a given occasion. Suppose that *s* is a system and *E* is a property realized by properties Q_1, Q_2, …. Then, "*s*'s having *E* on this occasion is identical with its having *Q* on this occasion. There is no fact of the matter about *s*'s having *E* on this occasion over and above *s*'s having *Q*. Each instance of *E*, therefore, is an instance of one of *E*'s

[1] In some formulations of the causal inheritance principle, Kim admits that the causal powers of the realized property may be a proper subset of those of the realizer. Kim's argument for identity does not go through with this assumption.

realizers, and all instances of E can be partitioned into Q_1-instances, Q_2-instances, … , where the Q's are E's realizers. Hence, the E-instances reduce to the Q_i-instances" (Kim 1999, 15-6).

Kim has long defended a "property exemplification" theory of events, according to which events are exemplifications by substances of properties at a time (Kim 1966, 1976). On this conception of events, an event is a "structured complex" and can be represented by an ordered triple, $\langle x, P, t\rangle$, where x is the constitutive object (n-tuple of objects) of the event, P is the constitutive property (n-adic relation) or "generic event," and t is the time when x has P (objects in the n-tuple bear the n-adic relation to one another). Given this conception, there are two plausible accounts of what a *property instance* is. It is either a token event, a property instantiation – x's having P at t – or the "token property" or "abstract particular" (i.e. trope) that is the particularization of the constitutive property of that event. Consequently, there are two interpretations of the token-identity claim regarding property instances.[2] Consider the token-identity claim as a claim about token Kim-type events. On this view, it is contradictory to assert property instance (event token) identity together with property (event type) distinctness. Suppose that one adopted the property

[2] One good example of authors talking past one another: Louise Antony assumes that what Kim means by "property instances" are the *entities* that have the property "as opposed to *tropes*" (Antony 1999a, 43 n.3). See also Antony (1999b), where she also adopts this view of property instances. But this is a flawed interpretation of Kim's argument and ontology. Interpreted as Antony does, the upshot of Kim's argument for the identity of property instances would be the unremarkable fact that system s is identical to itself – not a claim that one would think that Kim would need to argue for – since, as Kim makes clear, he thinks that *substances* (objects) *have* properties. One might be drawn to Antony's interpretation of Kim since it is presumably objects/substances that have causal powers, not tropes or "particularized properties." However, this also betrays a misunderstanding of Kim's ontology. For, he thinks that *events* (property instantiations/exemplifications) have causal powers (Kim 1999, 16). Antony's interpretation would make more sense if one adopted the view that events (i.e. concrete particulars) *have* properties. But this is not Kim's view, events do not *have* properties; they are partly "constituted" by properties, or they *exemplify* properties. So Antony is not agreeing with Kim but rather talking past him. Of course, both non-reductive and reductive physicalists will agree that every mental "property instance" is identical to some physical "property instance" or other. However, this amounts to claiming that the same subjects have mental and physical properties. This claim, coupled with the non-reductionist's denial of type-identity, makes it unclear how non-reductive physicalism differs from dualism. So the *informative* constraint is not satisfied.

exemplification view of events. The claim that every token mental event is identical to a token physical event amounts to the claim that every triple, $\langle x, M, t \rangle$, is identical to a triple $\langle x, P, t \rangle$, where M is a mental property and P is a physical property. But, given the individuation condition on events, this implies that the constitutive properties are identical, that the mental property M, is identical to the physical property P. If one adopts this view of events, token-identity implies type-identity.[3] So, formulating non-reductive physicalism as including the commitment to token identity implies that the *informative* constraint is violated; the ontological claims of non-reductive physicalism collapse into those of reductive physicalism. To avoid this, the non-reductive physicalist must reject type-identity for properties (or events). Consequently, she is forced to reject token-identity for properties (events) as well, if she adopts Kim's view of events. Without spelling out just how token mental events *are* related to physical events, non-reductive physicalism now collapses into dualism or emergentism, and the *informative* constraint is again violated.

Suppose that one endorsed Kim's theory of events but instead thought of property instances as tropes (particularized properties). Can the distinction be drawn on this view? Those who adopt this view and use it to develop non-reductive physicalism claim that every mental trope is identical to a physical trope but that these tropes can be grouped into distinct mental and physical property types (classes of resembling tropes) (see, e.g., (Robb 1997)). It is unclear that this is sufficient to make non-reductive physicalism interestingly different from reductive physicalism – that the *informative* constraint is satisfied. For, it is not clear that mental properties (trope types) are causally relevant or efficacious on this view. One could argue that, given that the effect was caused in virtue of the trope being physical, the fact that it was also

[3] As Sydney Shoemaker reminded me, Kim is aware of this problem and in response suggests that mental properties are not "constitutive properties of events" (Kim 1993b, 364-5 n.5). See also (Kim 1998, 56).

mental was entirely irrelevant (for a related charge, see, (McLaughlin 1993) and the references therein).[4] This would make mental properties epiphenomenal. It is also not clear that tropes are the kind of things that can be conceived of or grouped together in two different ways (i.e. as mental or as physical). This seems to make them too much like objects. Tropes are particulars, but they are *abstract* not *concrete* particulars. It seems that it is the concreteness of objects that allows them to lead a "hidden life" (to use Helen Steward's (1997) phrase) – to be conceived of under different descriptions or grouped in different ways. In any case, since the cogency of this approach turns on these issues about the nature of tropes, the *specificity* constraint is violated. Let's now turn to views of events according to which they are *concrete* particulars.

The main competing view of events, associated with W.V.O. Quine and Donald Davidson, sees them as concrete particulars – regions of space-time (or the contents of regions).[5] However, the claim of token-event identity, in this framework, is not strong enough for physicalism. Token-identity is now just the claim that there is

[4] We will see that this complaint is also levied at views of events that take them to be concrete particulars.

[5] Another instance of interpretative confusion: Cynthia and Graham MacDonald seem to adopt a Davidsonian view of events. However, in a footnote, they indicate that they believe that their view is in fact neutral between Kim's and Davidson's view of events (1986, 147 n.5). For, they claim that both Davidson and Kim "construe events as non-repeatable, dated particulars, and hence as individuals capable of possessing properties." But, as Antony does, they misrepresent Kim's ontology. Kim-type events do not possess properties in the way that Davidsonian events possess properties. A Davidsonian event has (or possesses) a property in the same way that a *substance* has (possesses) a property in Kim's ontology. In contrast, in Davidson's ontology, a fact or state of affairs exemplifies or instantiates a property in the same way that a Kimian *event* exemplifies or instantiates a property. So, the MacDonald's purported neutrality depends on conflating these two ways that properties can be related to entities.

According to the MacDonalds, events have properties by being instances of them. Since they claim that events are just "what's going on" in a region of space-time, they insist that events can have (instantiate) more than one property. They write: "Indeed, if an instance of a mental property just *is* an instance of a physical one, then, despite the distinctness of the properties themselves, the anomalous monist is right to insist that the former can be ... causally efficacious ... To insist otherwise, it seems to us, is tantamount to insisting that no event which is an instance of a mental property can be (i.e., be identical with) an instance of a physical one. And it is not clear that this is anything more than a dualist prejudice" (1986, 148-9). Far from being a dualist prejudice, as I showed above, it is an implication of the property-exemplification view of events. Once again, attempting to draw the distinction has led to disputes about the ontology of events, threatening to violate the *specificity* constraint.

a single region of space-time (or its contents) that has different properties. It does not require that the region causes the given physical effect in virtue of its mental properties rather than solely in virtue of its physical properties. Further, it does not require any tighter relation between mental and physical properties other than that they belong to the same regions and that the former supervenes on the latter.

So, this formulation also does not demonstrate how non-reductive physicalism is a distinct form of physicalism. On the Davidsonian view of events, one can compatibly assert token-identity while rejecting type-identity. But, asserting token event identity now amounts to claiming that the same region of space-time has two properties, a mental one and a physical one, and this does not show that it is the mental property (even the particular mental property instantiation *qua* mental) that is causally efficacious. Consequently, the *informative* constraint is not satisfied.[6] It is not clear that this characterization of non-reductive physicalism differs from reductive physicalism.

If one insists that causation is an extensional relation between Davidsonian events, then the exclusion problem is about the "quausation" relation (Horgan 1989) – causation in virtue of certain properties. Alternatively, one could reserve the term 'causation' for a relation between Kim-type events (or, if one objects to this view of events, insist that the causal relata are states of affairs or facts). These debates about the nature of events and causal relata only affect the way the mind-body problem and the distinction between non-reductive and reductive physicalists are formulated. As we have seen, a purely Davidsonian framework for events does not address all of the issues that are at stake – whether these issues are put in terms of whether Kim-type events with constitutive mental properties are ever causally efficacious, whether

[6] Latham (2003) also argues that token physicalism is not a coherent, substantive view that is stronger than minimal substance physicalism and weaker than property (type) physicalism.

mental states of affairs are ever causally efficacious, or whether there is mental "quausation."

For convenience, I assume a Kim-type account of events. However, given this framework, any non-reductive physicalist must deny both token- and type-identity for events. Thus, more must be said about the relation between mental and physical properties in order to distinguish non-reductive physicalism from dualism, emergentism, and other non-physicalist doctrines. I also assume a minimal causal theory of properties, as is customary in this debate. In the actual world, a single property always is associated with (or contributes) a unique set of causal powers, and no two properties contribute exactly the same causal powers (see (Shoemaker 2001), (Gillett 2002), (Kim 1998)).

1.2. Summary of the Dissertation

Chapter 2: The Exclusion Problem and the Causal Argument for Physicalism

The causal exclusion problem amounts to the claim that the two main theses of non-reductive physicalism are inconsistent. It can be presented as a set of four inconsistent propositions, each of which the non-reductive physicalist is apparently committed to: (I) *completeness of physics*: physical causes suffice for a complete causal account of the physical world; (II) *causal efficacy*: mental events are causally sufficient for physical effects; (III) *exclusion principle*: there cannot be two or more sufficient causes at a time of a given event (except for cases of overdetermination, which is not relevant here[7]); (IV) *irreducibility*: mental and other high-level properties are not identical to physical properties.

[7] At least if we use the term 'overdetermination' to cover only cases where two independent causal processes result in a single effect (e.g. firing squad cases). Some use 'overdetermination' more broadly to cover any case where there are two, numerically distinct sufficient causes of an effect at a single time.

There are many non-reductive physicalist responses to this argument, but I argue that they are all inadequate. One of the most popular responses is to note that the exclusion problem generalizes to all properties other than fundamental physical ones. Hence, something must be wrong with the reasoning that leads to the problem. Otherwise, one is led to the absurd view that either only fundamental microphysical properties are causally efficacious or no properties are causally efficacious (if there turns out to be no fundamental physical level). Although I think that the causal exclusion problem does generalize, this is not an adequate response to the exclusion problem. One needs to say exactly where the exclusionary reasoning goes wrong and offer a positive account of why fundamental physical sufficient causes do not always preempt mental and other broadly physical sufficient causes.

I suspect that the popularity of this "generalization argument" as a response to the exclusion problem, and the subsequent failure to state just which of theses (I)-(IV) is false, has led many non-reductive physicalists to overlook a tension in their position. That is, it is seldom explicitly stated that the causal exclusion problem is intimately related to one of the few arguments for physicalism: the so-called "causal argument," which uses theses (I)-(III) to argue for the denial of (IV): the claim that all (causally efficacious) properties are physical. Non-reductive physicalists have proffered versions of this argument, while at the same time claiming to maintain *irreducibility*. This blatant inconsistency is usually covered up with some gesture that seems to violate the strong exclusion principle just appealed to as a premise: for example, by claiming that mental properties need not be identical to physical properties but can be merely "congruent with" them – i.e., either type identical to, or realized by, physical properties (Papineau 1993, 21ff.).

The allegation that non-reductive physicalism is inconsistent can be put even more strongly once the relation between the exclusion problem and the causal

argument for physicalism is noted. Most non-reductive physicalist responses to the exclusion problem that go beyond the generalization argument claim that we should reject (III), the exclusion principle. Obviously, the conjunction of (II) and (IV), *causal efficacy* and *irreducibility*, is what is distinctive of non-reductive physicalism. And, giving up (I), the *completeness of physics*, would seem to amount to abandoning physicalism. However, simply rejecting (III) without endorsing *any* exclusion principle (one that would rule out *non-physical* properties that causally overdetermine physical events) deprives us of a sound argument for any form of physicalism. So these responses amount to throwing out the physicalist baby with the bathwater of reductionism. That is, they get stuck on the first horn of a dilemma that Jaegwon Kim has posed: either non-reductive physicalism collapses into a non-physicalist dualism (or epiphenomenalism), or it embraces the causal argument for physicalism and collapses into a reductive view. So any satisfactory response to the exclusion problem must show how to escape this dilemma by showing (a) why mental properties and their physical realizers do not compete for causal sufficiency and (b) why this does not undermine the causal argument for physicalism.

In the second chapter, I develop such a response. As many have noticed, two senses of 'physical property' have been used in the literature: a restricted-sense according to which, roughly, only the properties needed in an ideal fundamental physical theory count as physical, and a broad-sense according to which commonplace and special science properties like *being a giraffe, being slippery,* and *being a rock* count as physical properties.

I argue that we should also distinguish two senses of the "completeness of physics," which have sometimes been conflated in the literature: *fundamental force completeness* which is the claim that every fundamental force is physical, and *causal sufficiency completeness* which is the claim that every physical event which has a

sufficient cause at a given time has a physical sufficient cause at that time.[8] *Fundamental force completeness* together with the exclusion principle poses no problem to non-reductive physicalism because it does not rule out mental sufficient causes. This combination only rules out fundamental mental forces, which is common ground for all versions of physicalism.

In order for the exclusion problem to arise, the reductionist must claim that *causal sufficiency completeness* is minimally true of the restricted-sense physical domain: that every physical event has a restricted-sense physical sufficient cause.[9] I argue that even by the reductionist's own lights the *unqualified exclusion principle* (that rules out any numerically distinct simultaneous sufficient causes) is untenable. For, the reductionist must admit that there are many simultaneous restricted-sense physical events that are each causally sufficient for a given physical effect. The exclusion principle's plausibility derives from the general claim that there cannot be more than one independent causal chain or pathway leading to the same event (except for cases of "independent," firing-squad-like overdetermination, which do not apply to the mental/physical case) (Kim 1989, 243, n.15). Thus, a tenable exclusion principle will have to be qualified so that it does not result in competition for sufficiency between events that are part of the same causal chain or path.

I offer the following *qualified exclusion principle*: there can be two or more simultaneous sufficient causes of an event only if they are part of overlapping token causal processes or part of the same token causal process (barring cases of

---

[8] This claim is trivially true if the world is indeterministic, since in that case, no event would have a sufficient cause. I follow the common practice in the literature of bracketing questions about indeterminism. If causation is indeterministic, *causal sufficiency completeness* can be roughly formulated as the claim that physical antecedents suffice to fix the probability of every physical effect (cf. Papineau 1993, 2001).

[9] When discussing whether a set of entities is causally complete, one needs to identify the smallest set of physical entities which yields a true claim when substituted into the completeness schema, for short, the set of which the completeness claim is *minimally true*. Of course, the completeness claim would still be true of a larger set that included entities that are causally irrelevant, inefficacious or redundant, but these additions would be otiose and uninteresting. This is sometimes left tacit in what follows.

"independent" overdetermination). Hence, two events that figure in the same token causal process do not compete for causal sufficiency.[10] The many, apparently competing, simultaneous restricted-sense physical causes are part of the same causal process, so they do not compete for causal sufficiency. Likewise, I claim that a mental property and its physical *realizer* are part of the same token causal process, so they do not compete for causal sufficiency. In general, when property instantiation *Y* is realized by property instantiation *X*, there is a token causal process in which they both figure. Note that this constraint is not met by supervenience: if *Y* supervenes on *X* there need be no causal process in which they both figure. Indeed, there is no constraint that *Y* figures in *any* causal process. This provides another reason, in addition to problems raised by others for supervenience-based formulations of physicalism, why supervenience should not be used to formulate non-reductive physicalism.

Chapter 3: A Causal Argument for Physicalism and the Distinction between Reductive and Non-reductive Physicalism

Using *fundamental force completeness* together with the *unqualified exclusion principle*, and *causal sufficiency completeness* with the *qualified exclusion principle*, I provide a two-part causal argument for a form of physicalism that need not be reductive. As even some reductionists have pointed out, it is important that some properties of macrophysical objects be included in the domain to which *causal sufficiency completeness* applies; for example, masses of one kilogram have causal powers that no smaller masses have (cf. Kim 1998, 113-4). However, these powers are not fundamental but are determined by the causal powers of fundamental,

[10] Stephen Yablo's observation that determinables and determinates do not compete for causal sufficiency is a special case of my claim since determinables and determinates are plausibly part of the same token causal process (if either is part of any process). However, as I argue in Chapter 5, mental properties are *not* related to their physical realizers as determinables are related to their determinates.

microphysical entities. The empirical evidence that the conservation of energy holds universally in the actual world gives us reason to believe that the conjunction of *fundamental force completeness* and the *unqualified exclusion principle* are true of the actual world. This guarantees that non-fundamental powers do not involve any non-physical forces, in short, that *reducibility of forces* holds in the actual world. Further, *causal sufficiency completeness*, the *qualified exclusion principle*, and the causal efficacy of mental (and other broadly physical) properties provide an argument for *causal monism* – the thesis that no non-physical causes overdetermine physical effects. Together, these two arguments establish a form of physicalism that need not be reductive.

I then examine how to reformulate this argument for physicalism without the assumption that there is a fundamental level. I suggest that the question of whether there are any fundamental mental forces is a special instance of the question of whether mental entities are to be found below some level of decomposition. Both of these questions are concerned with whether mentality is a deep feature of the world on par with restricted-sense physical features like mass and charge, or, if it is rather a feature that appears only in certain very complicated systems and that is asymmetrically dependent on more basic physical properties. I offer *aggregational completeness*, roughly, the claim that mentality does not go "all the way down," as a generalization of *fundamental force completeness* and show how to use *aggregational completeness* as part of an argument for physicalism.

However, what I offer in Chapters 2 and 3 is merely a programmatic solution to the exclusion problem and defense of non-reductive physicalism. One still needs to develop an account of the relation between restricted-sense physical properties and mental properties. In order to be compatible with the qualified exclusion principle, all that is required is that mental properties and certain restricted-sense physical

properties are not independent. Physicalism requires the stronger claim that mental properties are determined by and dependent on restricted-sense physical properties (at least in the actual world). Further, non-reductive physicalism requires that this dependence relation allows mental properties to be distinct from restricted-sense physical properties and yet be natural, to be part of causal processes or mechanisms.

Further, an account of the determination relation between mental and restricted-sense physical properties is need to address further reductionist challenges regarding causal explanation or causation (as opposed to the relation of causal sufficiency). Even if mental and restricted-sense physical do not compete for causal sufficiency, they may be in competition with respect to *causation*, that is, they may compete for being the (or a) cause of a certain physical event. Further, restricted-sense physical properties may always win this competition. In Kim's earlier writings he appealed to a principle of causal-*explanatory* exclusion. As he refined the principle in response to criticism it became a principle about causal sufficiency. However, there are more reductionist worries lying behind the original causal-explanatory exclusion principle than are captured by the revised version in terms of causal sufficiency. For example, the early Kim claims that if one cause is dependent on or derivative from another, then the latter gains "explanatory or causal dominance over the [former]" (Kim 1989, 246). Further, he claims that restricted-sense physical causal accounts are always "deeper and more theoretical and systematic" (Kim 1989, 251) and "more detailed, more revealing, and theoretically more fecund" (Kim 1989, 249) than high-level causal accounts.

The second main line of reductionist attack is a way of posing these worries. Assume that some solution to the exclusion problem has been offered so that causal sufficiency does not drain away. The non-reductive physicalist claims that high-level properties are realized by, and hence dependent on, restricted-sense physical

properties. The second reductionist line claims that once an account of realization is spelled out one will see that the reason why high-level and restricted-sense properties do not compete for causal sufficiency is that high-level realized "properties" are merely shorthand for, or convenient descriptions that we use to refer to, the natural restricted-sense physical properties. It is only these restricted-sense physical properties that are really causally efficacious and do all of the causal work. Put another way, the claim is that there is no account of realization that allows realized properties to be a part of the causal structure of the world in the way that realizer properties are.

In Chapter 3, I argue for a way of drawing the distinction between non-reductive and reductive physicalism that applies whether physicalism is formulated in terms of identity, supervenience, or realization. I show how the two senses of "completeness of physics" correspond to two classes of natural properties: (i) the class of properties that is the minimal base for the determination of every causal power (but whose members do not contribute every causal power) – those of which *fundamental force completeness* is minimally true; (ii) the smallest class of properties whose members contribute all of the causal powers – those of which *causal sufficiency completeness* is minimally true. Reductionists have in effect claimed that these two classes are identical – that the fundamental physical properties can fulfill both *fundamental force completeness* and *causal sufficiency completeness* roles along with grounding similarity. Non-reductionists claim that (i) is a proper subset of (ii); hence, there are properties (those in (ii)/(i)) that fulfill the causality and similarity roles without being members of the smallest domain that satisfies the *fundamental force* completeness role. This shows why it is consistent to say that some properties are causally efficacious, natural, and irreducible in one good sense, but are *not* restricted-sense physical properties.

Thus, the debate between reductive and non-reductive physicalism is best understood as a debate about which properties are natural, not as a contest between token- and type-physicalism or between supervenience-based and identity-based formulations of physicalism. I end Chapter 3 with a discussion of mechanisms, which will be further developed in Chapters 5 and 6.

Chapter 4: Varieties of Realization

In Chapter 4, I use this debate about natural properties to explicate the difference between reductive and non-reductive interpretations of the supervenience relation that has been noted by others. I then argue that the non-reductive interpretation of supervenience cannot be used to formulate physicalism because it fails to satisfy what I call the *location* and *causal process* constraints. Supervenience fails to meet the location constraint roughly because it fails to elucidate the place of mental properties in a physical world; it does not require that mental properties are grounded only by restricted-sense physical properties. Supervenience fails to meet the causal process constraint in that it does not explain how mental properties and restricted-sense physical properties can be part of the same token causal process while remaining *distinct*.

I then investigate whether some account of realization fares better. I discuss Putnam's seminal papers on functionalism and explain how his use of the term 'realization' is multiply ambiguous. Failure to notice this has led to further confusion about the nature of the debate between reductive and non-reductive physicalism. According to one account of realization, the functional role account, realized properties are by definition second-order properties – the property of having some property or other that meets a certain condition or role. I show how this results in the realized "property" being merely another way of referring to the realizer. Consequently, the location constraint is met at the price of the causal process

constraint; mental properties and their physical functional role realizers are not distinct natural properties. To simultaneously meet these two constraints one needs a realization relation between two distinct natural properties. I discuss how two accounts, the subset account and the microrealization account, satisfy the location and causal process constraints when employed together. However, they seem to face a dilemma – to be caught between the inadequate physical grounding of mental properties and the reduction of those properties. In Chapter 5, I elaborate on some implicit assumptions made by these accounts to show how to avoid that dilemma.

Chapter 5: Realization, the Determinate/Determinable Relation, and Mechanisms

The various forms of the second line of attack on non-reductive physicalism differ in scope and in their reasons for thinking that realized properties are not genuine causes distinct from their realizers. Sometimes Kim claims that all realized properties are merely abundant properties, concepts, or property designators which refer to "first order" realizer properties (which may vary across species- or structure-types). In effect, he claims that what all of the things that satisfy a realized property have in common is something that is projected onto the world by us. Such realized property designators may be useful and *practically* indispensable since they group first order properties in ways that we humans find illuminating and helpful for communication (see Kim 1998, 105). However, realized properties do not mark out any real resemblances in the world; all of the objective similarity and causal efficacy is captured by their realizers.

At other times, Kim supports a similar conclusion by following a suggestion made by David Lewis that multiple realized properties are not eligible to be causes because they are too disjunctive. Other philosophers attempt to show that disjunctive properties whose causal features overlap in an appropriate way are natural. Hence, a multiply realized property can be reduced to the disjunctive property constructed from

all of its possible (non-disjunctive) realizers. Finally, some philosophers have argued that the requirements on when a property is genuinely multiply realizable are more demanding than is commonly assumed. They claim that mental properties are not in fact multiple realized and hence can be reduced to their unique realizers.

Stephen Yablo's (1992) proportionality constraint is a promising way of responding to the claim that realizers always preempt realized properties as causes:[11] sometimes realizers are not the cause of a given high-level effect because they are not proportionate to the effect; for example, they may include too much irrelevant detail, whereas realized properties are proportionate to the effect. However, Yablo is apparently committed to the view that mental properties are determinables of which their physical realizers are determinates, and he bases judgments of proportionality on considerations regarding determinables and determinates. I think that this gets the relation between mental properties and their physical realizers wrong and consequently misses the true metaphysical basis for proportionality and irreducibility.

I argue that mental properties are not related to their physical realizers as determinables are related to their determinates. I point out a crucial difference between paradigmatic instances of physical realization and instances of the determinable/determinate relation. Although having a determinate is a way of having a determinable and having a realizer is a way of having the property it realizes, the "way of having" is different in the two cases. This is suggested by the fact that there seems to be nothing to fill the blank in this analogy: to be in the physical condition K of this steaming tea is to be at 95°C in a certain *micromechanical* way just as to be the scarlet of Hester Prynne's letter A is to be red in a certain ___ way (see Yablo 1992, 253 for the tea example). While *being scarlet* is a way of *being red*, it makes little

[11] Sydney Shoemaker (2001) has developed an account of realization that can utilize the proportionality constraint.

sense to try to characterize further the kind of mechanism by which scarlet, as opposed to crimson, determines red.[12] By contrast, not only is a physical realizer a way of being the mental property it realizes, it is perfectly coherent to ask *how*, or by what kind of mechanism, the realizer determines the mental property.

I then spell out the metaphysical basis for this intuitive difference. Mental properties and their restricted-sense physical realizers belong to different aspect spaces, while determinables and determinates belong to the same aspect space. Restricted-sense physical realizers of mental properties will be complicated structural properties involving restricted-sense physical entities (e.g. cells, molecules, atoms, and ultimately subatomic particles or fields). Consequently, they will be characterized by restricted-sense physical aspects and distinguished along restricted-sense physical dimensions, e.g., physiological, chemical, and physical. That is, in order to specify how physical realizers of pain differ we will not say how they differ in their broadly physical (specifically, mental) aspects, e.g. in their "painness," in the aspects that characterize pains – for example, in the intensity, duration, affective or motivational aspects of the pain. Rather, we will list "non-pain" ways in which the realizers differ – different configurations of constituent entities, different charges, etc.

Relatedly, mental properties are abstract relative to their physical realizers in a different way than determinables are abstract relative to their determinates. This difference in abstractness is easily confused with the phenomenon of multiple realizability. A determinable is abstract relative to one of its determinates in that it takes up a larger volume of an aspect space common to both of them. A mental property is not abstract relative to one of its physical realizers in this way, but the way in which it *is* more abstract is harder to state precisely. If the intuitions behind the

[12] If one adopted the view that colors are surface spectral reflectances, then this claim might not be true. But the way in which determinables are multiply determined (if they are) is still different from the way in which realizers are multiply realized (if they are), as I discuss below.

multiple realization argument are correct, mental properties are more "modally flexible" or "compositionally plastic" (Boyd 1980) than physical ones. They are paradigms of the properties Robert Stalnaker describes as "more abstract, and so might apply to things even if the properties on which they seem to supervene did not" (1996, 233). They could be instantiated over a wider range of physical conditions or situations than their realizers could. I argue that even if realized properties are uniquely realizable they are abstract in that they are indifferent to which aspect space characterizes their realizers. For example, realizers of a mental property can be mechanical, neurochemical, hydraulic, etc. as long as these aspects "combine" or "aggregate" so as to result in the instantiation of the mental property. The restricted-sense physical aspects that characterize the realizer do not matter; the realized property is characterized by different aspects than its realizer even if it is uniquely realized.

This difference in abstractness is related to the fact that physical realizer types are often *multiply determinative* – they realize different kinds of properties, properties that fall under different determinables. By contrast, determinates of a given determinable are not multiply determinative. For instance, a determinate shade of blue determines only other less determinate colors.

These metaphysical differences correspond to the different ways that determinables/determinates and realized properties/realizers relate to *mechanisms*. Determinables and determinates both bear the same relation to what I call a *sustaining mechanism*: a mechanism that can be used to explain how a property is instantiated and persists through time. The same type of sustaining mechanism is responsible for the instantiation of both a determinate and its determinables. By contrast, a physical realizer *provides* (or is part of) a sustaining mechanism for the properties it realizes. This is reflected in the commonly used metaphor of realizers underlying or grounding

the properties they realize; determinates do not underlie or ground their determinables in this way. Further, the fact that physical realizers are multiply determinative is connected to a different sort of mechanism. The different kinds of properties that are simultaneously realized by a physical realizer correspond to different types of *integrative mechanisms* in which the structural realizer plays a role – where the integrative mechanism can be used to describe how the realized property is integrated into relations with properties realized by other restricted-sense physical properties. (For example, mean translational kinetic energy and other molecular properties provide an integrative mechanism for the relations between temperature, pressure, and other thermodynamic properties of gases.) By contrast, a determinate of a given determinable plays a role in only one type of integrative mechanism.

I conclude Chapter 5 by offering an account of realization, *structural realization*, which explicitly utilizes the notions of sustaining and integrative mechanisms. This provides a positive account of cases of subset realization that are not cases of the determinable/determinate relation and arguably allows *natural* restricted-sense physical structural properties to realize mental properties. I also show how microrealization does not result in the reduction of mental properties to natural restricted-sense physical states of affairs types (as suggested by the apparent dilemma developed at the end of Chapter 4). In short, the microphysical states of affairs types that appear in the account of microrealization are natural mental types, not natural restricted-sense physical types. This sets the stage for Chapter 6, where I argue that the natural restricted-sense physical structural realizers that appear in the account of structural realization cannot be identified with the mental properties they structurally realize. For, a given structural realizer will be multiply determinative, will simultaneously realize special science properties other than the mental property.

Chapter 6: Multiple Realizability, *Multiple Determinativity, and Irreducibility*

This mechanistic framework also provides a response to the attacks on multiple realizability's role in arguments for non-reductive physicalism. All of the forms of the second line of reductionist argument share a common structure. They assume that if a property is uniquely realized by a restricted-sense physical property, then it can be reduced to that unique realizer (whether it is a disjunctive property or a structural property common to a given species). Thus, they assume that multiple realizability is necessary for irreducibility. This assumption is also common in existing formulations of non-reductive physicalism. Consequently, they are vulnerable to recent reductionist arguments that question the significance and extent of multiple realizability.

In Chapter 6, I use the account of realization and integrative and sustaining mechanisms developed in Chapter 5 to show that multiple realizability is not necessary for irreducibility. Thus, the reductionist arguments fail. I then explain how multiple realizability and multiple determinativity are each facets of the real ground of irreducibility: a many-many relation between sustaining and integrative mechanisms. Multiple realizability is characterized by a single integrative mechanism corresponding to many sustaining mechanisms, while multiple determinativity is characterized by many integrative mechanisms corresponding to a single sustaining mechanism. I show how that this mechanistic account of realization provides a metaphysical basis for claims about the explanatory autonomy of the special sciences. This framework allows one to explain why realized properties are causes of (or causally explain) physical events and are not merely designators for low-level realizers. Since realizers are multiply determinative, citing them as a cause of some high-level event will include irrelevant detail; this is one reason why they will not be proportionate to the given broadly physical effect. Thus, multiple determinativity provides a metaphysical basis for claims about proportionality, which otherwise look

merely pragmatic or overly anthropocentric. Restricted-sense physical realizers do not carve up the world in every objective way: they do not capture all of the causal structure that exists. Finally, I show how multiple determinativity explicates some passages from two of the architects of the non-reductive physicalist consensus, Philip Kitcher and Hilary Putnam, which seem to suggest (contra the consensus they helped to establish) that multiple realizability is not necessary for irreducibility.

CHAPTER 2

THE CAUSAL EXCLUSION PROBLEM AND HEMPEL'S DILEMMA

Much of the literature on physicalism has focused on how to characterize the physical in order to ensure that physicalism is a substantive, possibly true thesis. The main impetus behind this project is to counter the claim that no such formulation exists – that "there is no question of physicalism" (Crane and Mellor 1990). In arguing for this claim, Tim Crane and D.H. Mellor provide a well-developed version of what has come to be called Hempel's dilemma, according to which physicalism is either false (if formulated in terms of current physical theory) or vacuous (if formulated in terms of ideal physical theory).[13]

As physicalists, non-reductive physicalists need to provide a response to Hempel's dilemma. But they also face a problem that does not affect reductive physicalists: the so-called causal exclusion problem, the upshot of which is that non-reductive physicalism is allegedly an inherently unstable position that collapses into reductive physicalism, eliminativism, or dualism.

These two topics – how to characterize the physical in order to formulate physicalism and how to solve the exclusion problem – have been treated independently of one another. This is surprising given that similar issues, particularly the completeness of physics, are involved in both areas of debate. I suspect that Hempel's dilemma and the exclusion problem have appeared intractable to non-reductive physicalists because they have not appreciated that an adequate response to each problem must be developed with an eye toward solving the other.

[13] For more on Hempel's dilemma, see, e.g. (Melnyk 1997) and the references therein.

In this chapter and the following one, I offer a unified non-reductive physicalist solution to Hempel's dilemma and the exclusion problem – a solution that provides the resources to formulate a two-part causal argument for a form of physicalism that need not be reductive. I argue that the standard response to Hempel's dilemma has little chance of succeeding because there is no single domain of entities that meets all of the constraints on physicalism. One of these constraints, that the physical domain be complete or exhaustive, is ambiguous, and no single set of physical entities is the smallest set that satisfies both senses of completeness. I then show how spelling out the two senses of completeness provides the material for an adequate solution to the exclusion problem – one that does not undermine a powerful, causal argument for physicalism, as, I claim, existing non-reductive solutions do.

2.1. Tensions in Characterizing the Physical: Hempel's Dilemma and Two Senses of 'Physical'

There are two constraints that any physicalist theory must meet. First, according to physicalism, the domain of physical entities must be complete. Roughly, the physical domain must provide for a complete inventory of empirical reality and a sufficient cause for every physical event. Physical theory must be true, complete, and exhaustive. This constraint develops out of the idea that physicalism and its ancestor, materialism, are monistic theories of the world. According to materialism, everything in the empirical world is ultimately made out of the same kind of stuff.[14]

Second, physicalism must be a "fundamentalist" doctrine in that physical reality must not include any fundamentally mental or vital entities. Even if mental

[14] It is also related to the doctrine of the unity of science which the logical empiricists endorsed and promulgated in a syntactic, reductionist form. According to the logical empiricists, the unity of science was to be revealed in the unity of scientific language; all scientific terms were to be defined in terms of a basic physical language. It is debatable whether non-reductive physicalists should reject the unity of science constraint altogether or merely the logical empiricist formulation of it (see (Cartwright 1999) and (Fodor 1974) for opposing views).

properties are physically acceptable, a subset of these physically acceptable entities, which does not include mental (or vital) entities, must be fundamental. If physicalism is true, then this proper subset grounds or determines all other entities, including mental ones, but not vice versa.[15] Roughly, this determination relation implies at least that once the distribution of entities in the proper subset of physical entities is fixed, then the distribution of all other entities, including mental ones, is fixed. Such a relation is required in order to exclude *sui generis* mental entities. Otherwise, physicalism is not an alternative to dualism.

The upshot of Hempel's dilemma is that these two constraints cannot be simultaneously met. Since abandoning relatively a priori materialist doctrines such as the view that everything is made of inert, impenetrable matter, physicalism has granted ontological authority to physical science. But a question immediately arises as to whether this is supposed to be current physics or an ideal physical theory. On one hand, if one claims that *current* physics provides an account of what it is for something to be physical, then the completeness constraint will not be met. Since every past physical theory has turned out to be false and incomplete, it is overwhelmingly likely that current physical theory is as well. If so, there are empirical properties that are undreamt of in current physics; hence, if one adopted this proposal, physical theory would be incomplete or inadequate.[16]

On the other hand, if we define physical entities as those referred to by an *ideal* physical theory, then the second constraint will not be satisfied. Physicalism will then

[15] It is debatable whether these "fundamentalist" aspirations can be met if there turns out to be no fundamental level (e.g. if matter is infinitely divisible). I discuss this further in the next chapter. See also, (Schaffer 2003a) and (Montero 2006).

[16] Recent evidence suggests that around 95% of the matter/energy in the universe is unaccounted for – is so-called dark matter and dark energy (or "quintessence"). This suggests just how inadequate the current Standard Model is as an inventory of the fundamental constituents of the universe. (For a summary of the evidence concerning the existence of dark energy, see (Rees 2003); for dark matter see (Baudis 2006).)

be either trivially or vacuously true (Crook and Gillett 2001, Loewer 2001, Papineau 2001) or indeterminate in content (Hellman 1985, 609). The most serious problem posed by this horn of the dilemma is that "physicalism" formulated in terms of ideal physics might turn out to be true when it clearly should be false. For example, even if fundamental mental properties or Cartesian souls turn out to be invoked by an ideal physical theory of empirical reality, "physicalism" would still be true according to this interpretation, which would clearly be unacceptable. In short, appealing to an ideal physical theory to formulate physicalism will not work because there is nothing to rule out entities that are clearly non-physical or physically unacceptable from figuring in that theory. Taking this horn amounts to extending the bounds of the physical beyond the breaking point. There is nothing to hold the content of physicalism relatively fixed and continuous with past physical theories and previous materialist doctrines. In addition, formulating physicalism in this way does not capture the claim that some physically acceptable properties are more basic than (ground or determine) others.

It is illuminating to note that Hempel's dilemma trades on a well-known tension in the vernacular conception of the physical. First, there are *contrastive* definitions of 'physical' like "involving the body as distinguished from the mind" or "pertaining to or connected with matter; material; opposed to psychical, mental, spiritual" (Oxford English Dictionary 1989). Opposed to these is an *expansive* understanding of 'physical,' which is almost synonymous with "natural": "concerned with natural laws and forces or material things" or "belonging or relating to Natural Philosophy or Natural Science; of, pertaining or relating to, or in accordance with, the regular processes or laws of nature" (Oxford English Dictionary 1989).

Unsurprisingly, there are two corresponding senses of 'physical' in the philosophical literature – what I will call a *restricted sense* and an *inclusive sense*. A classic example of the inclusive sense is Meehl and Sellars' definition, according to

which something is physical$_1$ if it "belongs to the space-time network" (1956, 252). In a similar vein is Herbet Feigl's gloss on physical$_1$ according to which it is "practically synonymous with 'scientific', i.e., with being an essential part of the coherent and adequate descriptive and explanatory account of the spatio-temporal-causal world" (1958, 377).

There are many ways to formulate a restricted sense of physicality, but they are all restricted in that they rule out (initially, or without a reductive argument) at least some special science entities (e.g. psychological, biological, or chemical ones) from the physical domain. One traditional way of formulating this sense of 'physical' is to say that physical entities are those that can be defined in terms of the vocabulary of fundamental physical theory (see, e.g. Hellman and Thompson 1975, 551). Feigl's physical$_2$ is also a restricted sense of 'physical': "the type of concepts and laws which suffice in principle for the explanation and prediction of inorganic processes" (1958, 377). Contemporary accounts of the restricted sense of 'physical' typically contrast "low-level physical" (Chalmers 1996, 33) or "quantum mechanical" (Sturgeon 1998) properties with "high-level physical" or "broadly physical" properties. "Low-level physical properties," according to Chalmers, are "the fundamental properties that are invoked by a completed theory of physics. Perhaps these will include mass, charge, spatiotemporal position; properties characterizing the distribution of various spatiotemporal fields, the exertion of various forces, and the form of various waves; and so on" (Chalmers 1996).[17,18]

[17] Note that properties like *mass* and *charge* are also possessed by macro-physical objects. Thus, they are not felicitously characterized as "microphysical" as Chalmers does. One might think we could capture the microphysical properties, if we restrict physical properties to those possessed *only* by fundamental entities (e.g. the quark colors, very small mass, etc.). Something like this is suggested by Frank Jackson's discussion of the physical. According to Jackson, the physical properties are those "that are needed to handle the non-sentient" or "those that we need to handle the relatively small" (1998, 7). However, if the fundamental physical entities turn out to be fields and all properties of non-sentient things are reducible to properties of fields, then the properties needed to handle the non-sentient will not be those that are needed to handle "the relatively small," as they will be properties that arguably apply to entities that extend through all (or large regions of) space-time. For these reasons, I prefer to

The standard response to Hempel's dilemma is to tinker with the definition of 'physical' in order to finesse a way between its horns. The goal is to come up with a single account of physicality that is not so broad as to make physicalism vacuous but not so narrow as to make it false or incomplete – to provide a single notion of the physical that meets both of the constraints introduced above. That is, the strategy is to attempt to identify a single domain of entities that is both complete (and thus provides a true theory of the world) and excludes mental entities (and thus is a genuine alternative to dualism).

Many recent attempts to do this lean toward the falsity horn and attempt to show that it can be blunted. Andrew Melnyk claims that we can define 'physical' in terms of the "entities and properties mentioned as such in the laws and theories of *current* physics" (Melnyk 1997, 623, italics added). He goes on to argue that even though a physicalist thesis based on current physics is likely false and incomplete, it only needs to be more likely than its relevant alternatives in order for it to be rational for us to take a "scientific realist" attitude toward it, as with any other empirical theory. Others claim that we can get an appropriate replacement notion for physicality simply by appealing to non-mental entities in one's formulation of physicalism (e.g., Spurrett and Papineau 1999).

This strategy takes the restricted-sense physical to be the only notion of physicality worth developing. If non-reductive physicalists accept this, then they are

---

use the phrase 'restricted-sense physical' instead of 'microphysical' even though the latter is more evocative.

[18] Many authors work under the assumption that the broad-sense physical and restricted-sense or "low level" physical domains are disjoint (e.g., Chalmers 1996, Sturgeon 1998). I will follow this convention and assume that the set of inclusive-sense physical properties is the union of the sets of broad-sense and restrictive-sense physical properties. Paradigm examples of broadly physical entities include properties such as slipperiness, wetness, and hardness; events such as handshakes; and objects such as giraffes, rocks and neurons. Note that the broadly physical domain *includes* mental properties if physicalism is true. The restricted-sense physical domain plausibly includes properties like spin and charge, objects like electrons and quarks, and events like the collapse of a quantum wave function and the excitation of an electron to a higher energy level in an atom. I am not assuming that there is a clear dividing line between broad-sense and restricted-sense physical entities, however.

committed to claiming that mental properties are *non-physical* (i.e. not restricted-sense physical). Accordingly, it would be incoherent to claim that mental properties *are* physical properties, even though this is apparently what all physicalists are committed to saying. Thus, non-reductive physicalism appears to be an inherently inconsistent doctrine.[19] Claiming that physicalism is true while at the same time insisting that mental properties are non-physical seems to require that the mental be brought into the physical fold – that the mental is "really something else" (as Fodor (1987, 97) remarks about intentionality), which seems to imply that every mental property must be paradoxically reduced to or identified with a physical property from which it is distinct.

Of course, the obvious way to make this position consistent is to opt for a reductive form of physicalism. The only alternative seems to be to admit that the mental is irreducibly *non*-physical after all, which seems tantamount to claiming that non-reductive physicalism is merely a contemporary version of emergentism or dualism. Non-reductive physicalism should avoid this false dichotomy and instead insist that mental properties *are* physical properties – that there is another, equally valid, notion of physicality, the inclusive-sense physical, which includes mental properties.

I think that the standard approach to Hempel's dilemma is misguided. There is little hope of arriving at one notion of 'physical' that will fulfill both of constraints – that is complete in every way and excludes mental properties.[20] For example, using only a version of the restricted sense of physicality to formulate physicalism (see, e.g. Melnyk 1997, Spurrett and Papineau 1999) obviously succeeds in meeting the second constraint – excluding mental entities (and other entities that would clearly be

[19] Cf. the seemingly inconsistent versions of non-reductive physicalism discussed below.
[20] I am not claiming that all existing responses to Hempel's dilemma fail, only those that try to make due with a single notion of physicality.

incompatible with physicalism, such as fundamental vital properties). But, as I argue below, the set of restricted-sense physical properties is unable to meet this constraint while at the same time satisfying the first constraint – providing for a domain of properties that is complete in every relevant way. This is not merely because it is likely that the entities invoked by current physical theory do not provide a true and complete account of physical reality. Rather, it is because there are two senses in which a set of entities can be complete, and there is no single set of physical entities that is the smallest set of which both senses of completeness are true.

Likewise, formulating physicalism with only a version of the inclusive sense of physicality (as (Poland 1994) does by appealing to future or ideal physical theory) provides for a true and complete theory of the world, but by itself it does not provide the means to capture the fact that a proper subset of these inclusive-sense physical properties is fundamental (in a way to be specified below). That is, a physicalist doctrine stated using only the inclusive sense of physicality is not by itself a substantive thesis – e.g., it is compatible with emergentism and dualism (without any additional claims or provisos). For example, Meehl and Sellars' and Feigl's definitions of 'physical$_1$,' if used to formulate physicalism, both threaten to render physicalism vacuous, since fundamental mental phenomena, if there were any, would arguably be "scientific" and would belong to the "space-time network."[21] This strategy easily avoids the falsity horn of Hempel's dilemma but has difficulty with the vacuity/triviality horn.

As the rest of this chapter and the beginning of the next make clear, my strategy is to embrace the vacuity/triviality horn of the dilemma with respect to causal

[21] As Nick Sturgeon pointed out to me, these definitions of 'physical' would not make physicalism completely vacuous, for such a physicalism would not encompass supernatural phenomenon that are outside of space-time. Nevertheless, any "physicalism" formulated solely in terms of them would not have much substance.

sufficiency but argue that this does not automatically undermine physicalism. I argue that one type of completeness, what I call *causal sufficiency completeness*, is minimally true of a domain that includes at least some complex aggregates of maximally restricted-sense physical entities.[22] I then argue that the fact that aggregates must be included in the smallest domain of which *causal sufficiency completeness* is true entails that an *unqualified exclusion principle* regarding sufficient causes is too strong. Without any further claims about the nature of aggregation, this is compatible with fundamental vital or mental forces or causal powers appearing in those aggregates. I claim that we can prevent physicalism from becoming a vacuous or trivial thesis by recognizing that another type of completeness, *fundamental force completeness*, if it is true at all, is minimally true of the set of maximally restricted-sense entities (taken individually). However, any *unqualified exclusion principle*, when combined with *fundamental force completeness*, will pose no unpalatable dilemma for non-reductive physicalists. I turn now to a discussion of these two types of completeness.

2.2. Ambiguity in the Claim that Physics is Complete

The standard approach to Hempel's dilemma fails because it overlooks the fact that there are two theses that have been expressed by the phrase 'completeness of physics.' As I argue below, both theses must be endorsed in order to make physicalism a substantive thesis, but they are not both *minimally true* of the same set of entities – that is, there is no single set of physical entities that includes all *and only*

[22] I adopt the term 'aggregation' from Kim (see quoted passage below on p. 38) and use it to refer, in the first place, to aggregate objects and also to aggregate events – events whose constitutive object is an aggregate. The constitutive properties of these events will also be, in a sense, aggregate properties. As I discuss in chapter 4, a property instantiated in an aggregate object will be microrealized by a microphysical states of affairs that consists of the constituents of that object propertied and related in certain ways.

the entities needed for the truth of the two completeness claims.[23] Hempel's dilemma helpfully points out the trouble that results when both senses of completeness are collapsed into a single thesis: one ends up with an unsatisfactory formulation of physicalism because any given set of physical entities is not the smallest set of which *both* completeness claims are true. Put another way, I think that Hempel's dilemma shows that physicalists need to appeal to two sets of physical entities when formulating physicalism.

The ambiguity in the completeness of physics shows up in a disagreement among physicalists about the minimal set of physical entities for which completeness holds – whether it is the set of fundamental microphysical (i.e. maximally restricted-sense physical) entities (assuming there is such a fundamental level[24]) or some more inclusive set of physical entities.

Most physicalists seem to think that the set of fundamental microphysical entities is causally closed. In a recent paper, Kim writes: "It is only when we reach the fundamental level of microphysics that we are likely to get a causally closed domain" (2003, 173). Scott Sturgeon (1998, 415) claims that completeness holds only for the quantum mechanical domain, and Lynne Rudder Baker claims: "[T]he notion of causal closure of the physical applies only at the lowest level of micro-physics" (1993, 79). Jonathan Schaffer also assumes the "completeness and closure of micro-causality" (2003b, 25). Finally, David Papineau supposes that fundamental restricted-sense physical entities will suffice for completeness: "Current physics, I take it, aims

[23] To anticipate: *fundamental force completeness* is minimally true of the maximally restricted-sense physical domain, but *causal sufficiency completeness* is minimally true of a more inclusive physical domain. *Fundamental force completeness* is still true, even if we quantify over inclusive-sense physical entities (all fundamental forces are inclusive-sense physical simply because they are all restricted-sense physical, and the restricted-sense physical is a subset of the inclusive-sense physical). However, broad-sense physical entities are irrelevant for the truth of *fundamental force completeness*; they are not needed to get a complete inventory of every fundamental force.

[24] In Chapter 3, I discuss some of the complications that arise when this assumption is dropped.

to develop a complete theory of paradigm physical effects in terms of the categories of energy, field and spacetime structure" (1993, 31). While he allows that the physical categories may need to be supplemented, he does not believe they will need to be supplemented by psychological categories (or, presumably, biological, geological or any other broadly physical categories). While these authors may hold different views on the class of micro-physical entities, they all claim that completeness applies to the physical domain only if 'physical' is taken in a *maximally* restricted sense, the level of basic microphysics.

In his 1998 book, however, Kim takes the opposite view from the one put forth in his 2003 paper. Here he asserts that completeness is not at all plausible if we construe 'physical' as 'microphysical':

> Perhaps the standard micro-macro hierarchy encourages the idea that the causally closed physical domain includes only the basic particles and their properties and relations. But that is a groundless assumption. Plainly the physical domain must also include aggregates of basic particles, aggregates of these aggregates, and so on, without end; atoms, molecules, cells, tables, planets, computers, biological organisms, and all the rest must be, without question, part of the physical domain. … On this understanding, being a water molecule is a physical property, and being composed of water molecules (that is, being water) is also a physical property. It is important that these micro-based properties are counted as physical, for otherwise the physical domain won't be causally closed. Having a mass of one kilogram has causal powers that no smaller masses have, and water molecules, or the property of being water, have causal powers not had by individual hydrogen and oxygen atoms. (Kim 1998, 113-4)

Note that that the properties that Kim here calls "micro-based" (and I call "micro-structural" or "structural") belong only to aggregates. In general, a structural property is the property of having parts with certain properties that are related in certain ways. So, in this passage, completeness is asserted to apply to the physical

realm only if 'physical' is construed in a fairly *inclusive* sense, that includes some aggregate entities and their structural properties.

2.2.1. Fundamental Force Completeness

I think this disagreement is best explained by the fact that the two sides are talking about different varieties of completeness. The claim that the maximally restricted-sense physical domain is the minimal domain that is causally complete is best interpreted as a claim about fundamental forces, at least if we take the maximally restricted-sense physical domain to include only individual restricted-sense entities and those forces which they individually contribute.

> *Fundamental force completeness*: all fundamental forces are physical. More precisely, fundamental physical forces are sufficient for the causal processes leading to every physical event.

When the microphysical domain is substituted into *fundamental force completeness*, it results in the claim that the existence of only microphysical entities is sufficient for a complete set of fundamental forces.

This sense of completeness is present in the literature but has not received much attention (especially in connection with the exclusion problem). It is clearly suggested by passages in Papineau's work (2001, 2002b, appendix), where completeness is taken to rule out the existence of *sui generis*, that is, fundamental, non-physical forces—for example, mental or vital ones. To deny the completeness of physics, Papineau claims "would in effect postulate an extra mental force alongside the fundamental physical forces of gravity, the electroweak force, and the strong nuclear force. This might once have made sense, but the cumulative evidence of two centuries of physiological research weighs heavily against it" (1996, 4). Jessica Wilson (2001) has recently defended a formulation of physicalism that relies on what amounts to a completeness claim about fundamental physical forces. Finally, one of Horgan's formulations of completeness is also along these lines. He claims that the

causal completeness of physics "means that non-physical properties cannot be causally basic properties—ones that generate fundamental forces that combine with physical forces to yield net forces different from the net resultants of physical forces" (1993, 573).

2.2.2. Causal Sufficiency Completeness

As the passage just quoted from Kim 1998 points out, it is arguably not true that the maximally restricted-sense physical domain is causally complete in another sense. That is, the maximally restricted-sense physical domain does not include all of the entities that are needed to provide a sufficient cause for any given physical event. This sense of completeness, what I call *causal sufficiency completeness*, is the one that has received the most attention in the literature (see, e.g., Lowe 2000, Montero 2003) and is the one invoked in presentations of the exclusion problem.

> *Causal sufficiency completeness*: Every physical effect is completely causally determined at time *t*, insofar as it is causally determined, by physical causes that occur at *t*. In other words, for every physical effect and every time *t*, if that effect has a sufficient cause that occurs at *t*, then it has a physical sufficient cause that occurs at *t*.

Causal sufficiency completeness amounts to the claim there is no need to go outside of the physical domain in order to find a sufficient cause for every physical event. This is usually thought to leave open the question of whether non-physical events can be sufficient causes of physical events. While they may not be needed in order to order to find a sufficient cause for every physical event, they are not ruled out.[25]

[25] However, one might wonder whether the existence of non-physical causes of physical events is in fact compatible with the completeness of the physical domain. Sometimes causal sufficiency completeness is taken to be the claim that the non-physical and physical cannot causally interact, that there can be only physical causes of physical events. (This is sometimes put as the "causal closure" of the physical domain, e.g. Kim 1998, 40.) This thesis, what I will call "strong completeness" (following Kim 2003, who calls it strong closure), has not always been clearly distinguished from causal sufficiency completeness as articulated above. If strong completeness is used in the exclusion problem, then a separate exclusion principle is not needed. I think that the exclusion problem is clearer and more convincing if we keep completeness and exclusion as separate premises.

We can think of *fundamental force completeness* and *causal sufficiency completeness* as schemas into which different sets of physical entities (corresponding to different notions of physicality) can be substituted. Assuming that the same set of physical entities (say, the set of basic microphysical ones) is substituted into the two schemas, *causal sufficiency completeness* is logically stronger than *fundamental force completeness*. The former entails but is not entailed by the latter. Assume that *causal sufficiency completeness* is true of the set of microphysical entities. Then, there cannot be any non-microphysical forces[26] that enter into the chain of causes that determines any given physical effect. If there were, then such a chain would not be completely microphysical, contradicting the assumption that *causal sufficiency completeness* is true of that domain. Hence, *fundamental force completeness* is true of the set of microphysical entities. On the other hand, assume that *fundamental force completeness* is true of the microphysical domain. The passage from Kim presents reasons for thinking that *causal sufficiency completeness* need not be true of the microphysical domain. For, a macrophysical whole could possess properties that were not possessed by any of its parts but were needed to provide a sufficient cause a given physical effect.[27] So, *fundamental force completeness* does not entail *causal sufficiency completeness*.

As I interpret the 1998 passage, Kim is claiming that we need to enlarge the set of physical entities beyond the basic microphysical entities and their properties in order to arrive at the minimal domain of which *causal sufficiency completeness* is true.

---

[26] i.e. forces that are contributed only by non-microphysical entities

[27] These can arguably include only "weakly emergent" properties (i.e. properties had by wholes but not by their parts), like the transparency of glass, which are compatible with physicalism (Cf. Kim: "A neural assembly consisting of many thousands of neurons will have properties whose causal powers go beyond the causal powers of the properties of its constituent neurons, or subassemblies, and human beings have causal powers that none of our individual organs have" (1998, 85)). *Fundamental force completeness* is not compatible with "strongly emergent" forces – forces that require the invocation of novel fundamental fields or laws, perhaps, e.g., the *élan vital.*

I agree, but I believe that the way in which it must be enlarged (by including at least some aggregates and their properties) implies that the *unqualified exclusion principle* is too strong, even by reductionist lights.

2.3. The Exclusion Problem: A First Pass

The exclusion problem can be set up in terms of four mutually inconsistent propositions:

> *Completeness of Physics*, which I take to be *causal sufficiency completeness*, as is standard in the literature (more on this below). For every physical event and every time *t*, if that physical event has a sufficient cause at *t*, then it has a physical sufficient cause at *t*.
>
> *Causal Efficacy*: Some mental events are causally sufficient for some physical events.
>
> *Unqualified exclusion principle*, which will be discussed further below, but for now can be taken as the claim that (EPa) there cannot be two or more sufficient causes of an effect at a given time, except for cases of overdetermination, and (EPb) there is no systematic overdetermination in cases of mental causation.[28]
>
> *Irreducibility*: Mental properties are distinct from (i.e. not identical or reducible to) the physical properties upon which they supervene (or which realize them).[29]

Jaegwon Kim, the main architect of the exclusion problem, claims that its upshot is that physicalists about the mind must choose between giving up

---

[28] The specific formulation of the exclusion principle has changed over the years (see, e.g., Goldman 1969, Kim 1989, 2003). I follow Kim's recent practice of using causal sufficiency (instead of causation or causal explanation), although, as I discuss in Chapter 3, the reductive physicalist may also think there is an exclusion problem with respect to these other relations.

[29] This formulation owes a lot to Crane (1995), Sturgeon (1998), Papineau (2001), Horgan (1997) and several of Kim's works (1989, 1998, 2003). As noted in Chapter 1, I am assuming a Kim-style account of events, according to which they are property instantiations represented by ordered triples of a constitutive object, a constitutive property, and a time. More broad-grained, Davidsonian accounts of events (and token identity theories stated in terms of them) do not address all of the worries raised by the exclusion problem (see, e.g., McLaughlin 1993). If one rejects a Kim-style account of events on the grounds that it is actually an account facts or states of affairs, I am happy to claim that the causal relata are facts/states of affairs instead of events (whatever those might be). Alternatively, one could insist that causation is an extensional relation between Davidsonian events. Under this assumption, the exclusion problem is about mental "quausation" (see Horgan 1989).

irreducibility, by endorsing reductive identities, or giving up *causal efficacy*, by admitting that the mental is causally impotent (Kim 2003, 165-6). Denying the *completeness of physics* is supposed to be tantamount to abandoning physicalism, and the *exclusion principle*, is alleged to be a metaphysical principle supported by sound arguments (see, e.g., Kim 1989). Sometimes Kim presents the exclusion problem as an argument that takes *completeness of physics*, *irreducibility* and the *exclusion principle* as premises and the denial of *causal efficacy* as its conclusion.[30] The purpose of this argument is to show that non-reductive physicalism is unstable and ultimately to support "'causal physical reductionism,' the thesis that any mental property—in fact, a property of any kind—that is causally efficacious must be a physical property or be reducible to physical properties" (Kim 2003, 152). Thus, Kim thinks that the non-reductive physicalist is forced to deny that irreducible mental properties are causally efficacious. And, ultimately, he thinks that instead of accepting this epiphenomenalist conclusion one should instead give up *irreducibility*.[31]

2.3.1. Tensions in the Non-Reductive Physicalist Interpretation of the Exclusion Principle

While the reductionist response to the exclusion problem is straightforward – "either reduction or causal impotence" (Kim 2003, 165) – the non-reductionist response is more complicated. Many non-reductive physicalists seem to be deeply ambivalent about how to respond to the exclusion problem. There are scores of papers that attempt to show how to defuse the exclusion problem, usually by arguing that the

[30] In Kim's recent work (1998, 2003, 2005), he takes this as one horn of a dilemmatic argument which he calls the "supervenience argument." Kim thinks of his argument for conditional reductionism: if mental properties are not identical to physical properties, then they are causally inefficacious.

[31] David Lewis (1966) puts forth an argument with (I) and (II) as explicit premises and (IV) as a conclusion. Lewis's version of the argument can be interpreted as assuming that (I) is a "strong completeness" claim, which incorporates (III). (Lewis rejects the causal equivalence of a physical state and its epiphenomenal correlate (1966, 25), in order to get asymmetry between the physical and the mental.)

exclusion principle is false as stated or showing that it leads to a *reductio ad absurdum* (the so-called "generalization argument" which I discuss below). However, many non-reductive physicalists also endorse an argument for physicalism – the so-called "overdetermination argument" (Sturgeon 1998) or "causal argument" (Papineau 2001), as I will call it – that has *completeness of physics*, *causal efficacy* and the *exclusion principle* as premises and (at least something very similar to) the denial of *irreducibility* as its conclusion. It is not clear how these two facets of non-reductive physicalism can be made consistent with one another. For, the first position denies the *exclusion principle* (or perhaps the *completeness of physics*),[32] while the causal argument *uses* that very claim as part of an argument for the denial of *irreducibility*. Thus, endorsing the causal argument for physicalism while adopting the standard response to the exclusion problem just trades one inconsistency (accepting the four theses that make up the exclusion problem) for another (accepting an argument that uses the *exclusion principle* as a premise, while at the same time denying that the *exclusion principle* is true).

Of course, there are other arguments for physicalism – e.g., arguments from the best explanation (e.g. Hill 1991) and from simplicity (e.g. Smart 1959). According to the argument from simplicity, identifying mental states with physical ones results in a simpler, neater ontology by getting rid of unreduced mental states and psychophysical laws which would otherwise be "nomological danglers" (i.e. states and laws that would not fit into the law-governed physical universe, but would be extraneous add-ons). According to the explanatory argument, type identities between mental and physical properties/events are the best explanation of the mental-physical

[32] Non-reductive physicalist obviously cannot reject *causal efficacy* or *irreducibility*. To do so would be to acquiesce in the reduction or epiphenomalism dilemma that supports Kim's conditional reductionism.

correlations observed in nature. This is supposed to provide a good and sufficient reason to believe that type physicalism is true.

However, it is unclear if these arguments are successful. The straightforward simplicity argument seems to be question begging – the question of whether positing a separate domain of mental entities is needed to save the phenomena is precisely what is at issue between physicalists and dualists. There is no neutral ground regarding the phenomena that must be explained by the candidate theories from which the "simplest" is chosen. For example, should we include the possibility of zombies and inverted qualia as among the phenomena that must be explained by the best theory? (Cf. Kim 2005. For a more elaborate criticism of the argument from simplicity, see Hill 1991, 26-40).

Turning to the explanatory argument, one reason to doubt its efficacy is that it seems to assume that the only relevant explananda are psychophysical correlations. A better abductive argument for physicalism would claim that it offers the best explanation of all of the relevant phenomena. And, as just noted above, which phenomena need to be explained will itself be controversial. In any case, if an explanatory argument (that is distinct from the causal argument) works at all, it works as an argument for reductive physicalism. For, it is less plausible to claim that a supervenience or realization relation, and not identity, is the best explanation of the psychophysical correlations. One needs additional arguments for the claim that one of these relations, and not identity, holds between the mental and the physical. Put another way, the physicalist needs to give an account of the relevant supervenience and realization relations and explain why they are to be preferred over identities.

Unless another argument for physicalism can be found, the non-reductive physicalist needs a version of the causal argument in order to provide some reason to believe in physicalism. In adopting the view that a causal-exclusionary argument for

physicalism may leave open whether or not *non-reductive* physicalism is true, I am in disagreement with those non-reductive physicalists who believe that the "underlying spirit" of the causal exclusion problem (and, hence, the causal argument) is utterly misguided. For instance, Terence Horgan claims that "if causal-exclusionary reasoning is sound at all, then what it really shows is that [non-reductive physicalism] is just mistaken" (1997, 171). Horgan apparently thinks that no non-reductive physicalist position is compatible with a broadly causal-exclusionary argument for physicalism. I think that such a combination of views had better be consistent. For, without a version of the causal argument, we have no good reason to think that any form of physicalism is true.

Yet, as mentioned above, existing attempts by non-reductive physicalists to endorse a version of the causal argument are unsuccessful. To take one example: David Papineau has developed versions of the causal argument since the mid-1990s (1993, 2001). The causal efficacy of the mental, together with the completeness of physics and an assumption that overdetermination is not widespread, implies that "we need somehow to identify the mental cause with the physical cause, so as to avoid the conclusion that [the physical effect in question] was overdetermined" (Papineau 1993, 23). Papineau thinks that we can avoid type-identity and token-identity of mental and physical properties by appealing to a "generous" conception of causation according to which "an instance of a strongly supervening fact causes the effects of those facts on which it supervenes" (Papineau 2002a, [online revision of his 1993]). In effect, this is the "supervenient causation" proposal put forward by Kim in his (1984). As Kim later realized, such a move is unsatisfactory. If there can be two distinct simultaneous sufficient causes, then the exclusion principle is violated. Merely claiming that two distinct events are "somehow identical" or "congruent," without spelling out in detail what this means, is unsatisfactory (see Papineau 1993, 21ff.).

Thus, the causal argument as stated in the literature leads to a reductive form of physicalism.[33] If one wishes to avoid a reductive form of physicalism, as Papineau seems to, then one must appeal to a weaker, qualified version of the exclusion principle. However, it is not clear what this qualified principle will look like, nor is it clear if it can be motivated independently of the non-reductive physicalist conclusion it is supposed to support.

Attempts to endorse the causal argument while distinguishing the resulting position from reductive versions of physicalism have led to some odd, apparently contradictory, formulations of physicalism. For instance, the conclusion of one version of Papineau's causal argument for physicalism is that every mental event is identical to a non-mental event (2001), a conclusion that many philosophers would like to avoid. Terence Horgan's version of non-reductive physicalism is also strange on its face: "mental causation via non-physical properties can co-exist with physical causation even if the physical realm is causally closed" (Horgan 1997, 166). Here, the non-reductive physicalist claims that the mental is non-physical, and causation by non-physical properties is alleged to be compatible with physicalism.

This ambivalence about how to deal with the exclusion problem and the resulting odd formulations of non-reductive physicalism can be explained by a tension in the non-reductionist's interpretation of the exclusion principle. Non-reductive physicalists seem to want to have things both ways. On the one hand, the non-reductive physicalist needs a strong version of the exclusion principle to argue for physicalism and rule out spooky, non-physical entities as causes. On the other hand, in order to maintain the non-reductive portion of her thesis, she needs a weaker version of the exclusion principle that is *not* violated if two "intimately related"

[33] Either (possibly species-restricted) type-identities or token-identities between Kim-events that are incompatible with the distinctness of mental and physical properties.

properties are simultaneously instantiated in the same object and are causally sufficient for the same effect. However, this claim – that there is some intimate relation between mental and physical events that is incompatible with dualism but such that it does *not* result in (an implausible kind of) overdetermination is perilously close to what the non-reductive physicalist wanted to argue for in the first place. This tension is different from the one that Kim emphasizes. Non-reductive physicalists do not simply seem to face a dilemma between reductionism and epiphenomenalism. They also seem to face a dilemma between endorsing a valid argument for reductive physicalism and endorsing an ineffective argument for non-reductive physicalism, one that is ad hoc, question begging, or at best incomplete (in that the resulting conclusion is weaker than physicalism, merely a kind of causal *monism*).

Existing versions of non-reductive physicalism face the following problem. If a strong version of exclusion is used, it is not clear how a reductive identity theory can be avoided. If a weaker version of the exclusion principle is used, it is not clear how it can be supported by arguments that are independent of the desired non-reductive physicalist conclusion. That is, it is not clear that there is an independently motivated, non-ad-hoc exclusion principle that will rule out ubiquitous, pernicious overdetermination, like firing squad cases, while allowing innocent (or pseudo-) overdetermination by mental properties and restricted-sense physical properties. (More on overdetermination below.)

Thus, an adequate non-reductive physicalist response to the exclusion problem must not merely reject the exclusion principle but also offer a replacement that can be used in an (hopefully sound) argument for a version of physicalism that need not be reductive. I claim that the two different varieties of causal completeness and corresponding interpretive tensions regarding the exclusion principle allow for this kind of response – one that claims that the trouble with the exclusion problem lies not

with any one of its theses taken on its own, but rather with how the completeness of physics and the exclusion principle interact.

I argue that *fundamental force completeness* and *causal sufficiency completeness* interact differently with a strong, unqualified exclusion principle and a weaker, qualified exclusion principle. *Causal sufficiency completeness* would rule out mental causes when combined with the unqualified exclusion principle. However, it is minimally true of a fairly inclusive set of physical properties which *already includes* many complex biological, chemical, and macrophysical properties. In this fairly inclusive domain, even reductionists can only consistently endorse a *qualified* version of the exclusion principle, which results in no competition between simultaneous causes that are related in the right way. As I argue in Chapter 3, an *unqualified exclusion principle* is still needed as part of the causal argument for physicalism, but such a principle is cotenable only with the weaker claim of *fundamental force completeness*, which also poses no threat to unreduced mental causation. In Chapter 3, I use these resources to develop a novel, two-part causal argument for a form of physicalism that need not be reductive. In the rest of the dissertation, I argue that non-reductive physicalism should be preferred over reductive versions.

2.3.2. Existing Non-Reductive Physicalist Responses to the Exclusion Problem

One common non-reductive physicalist response to the exclusion problem is to deny the exclusion principle – to claim that (EPa) is false, that there can be more than one sufficient cause of an event at a given time, *without* there being overdetermination.[34] This is often called a "compatibilist" response, following one of its major proponents, Terence Horgan (e.g., 1993, 573; 1997, 171; see also Bennett (2003), Block (2003, 135, 149), Yablo (1992), and Thomasson (1998) who asserts

[34] At least without something like the classic cases of overdetermination, like a death caused by two bullets striking simultaneously, where each overdetermining cause follows a separate and independent causal pathway.

further that a compatibilist, layered view cannot countenance causal interaction between layers). A small, but growing, number of authors challenges (EPb), the second part of the exclusion principle, and claim that widespread overdetermination is not as bad or implausible as Kim and others assume it to be. These authors claim that overdetermination is, in fact, ubiquitous and that it allows us to make sense of causation by macroscopic objects, properties, and events in general (see, e.g. Schaffer (2003b), Sider (2003)). For example, according to such a proposal, events involving some micro-particles and a simultaneous macro-event composed of those micro-events are each causally sufficient for a given effect, such as a window breaking.

It is not clear that there is much difference between these two types of response in the end. Those who deny (EPa) are merely claiming that mental causation is not like the classic cases of overdetermination – e.g., a death caused by the simultaneous impact of two bullets (each of which would have been fatal alone) or a window breaking by being simultaneously struck by two rocks (each of which would have broke the window alone). They still hold that mental events and their physical bases are numerically distinct and each is a sufficient cause of a given effect. Those who deny (EPb) think of overdetermination more broadly, and use it to refer to any case where we have a sufficient cause that can be "decomposed" into two sufficient causes.[35] So those who formulate their response as a denial of (EPa) agree that this broader kind of "overdetermination" is widespread and that mental causation is an instance of it. They may insist that it should not be called 'overdetermination' due to the fact that the mental/physical case arguably does not involve conspiracy or coincidence; two independent token causal processes do not result in a single effect, as they do in firing squad examples. But this is merely a terminological dispute. I will

[35] As in cases of "quantitative overdetermination," as when a boulder breaks a window and its two hemispheres "overdetermine" the breaking (see Schaffer 2003b).

use the term 'overdetermination' in the broad sense to refer any case in which there are two simultaneous sufficient causes of a given event.

Those who adopt this kind of response need to distinguish between innocuous and pernicious (firing-squad-like) kinds of overdetermination in such a way that any purported overdetermination by physical events and simultaneous, fundamental *sui generis* mental events falls into the pernicious category ("pernicious" because they would be improbable if as widespread as alleged cases of mental causation). Otherwise, the dualist could simply avail herself of the same general strategy and claim that overdetermination by non-physical mental events and physical ones is as ubiquitous and unproblematic as the kind of overdetermination that occurs when micro-events and a macro-event composed of those micro-events are both sufficient causes of some macrophysical effect. Further, one who pursues this kind of response cannot characterize the innocuous cases of overdetermination as those in which the simultaneous sufficient causes are related by whatever relation the non-reductive physicalist claims holds between mental and physical events (e.g. supervenience or realization). To do so would make the proposed response to the exclusion problem question begging or, at least, ad hoc. It would amount to the claim that overdetermination is not allowed (or would be an implausible coincidence) except in cases of mental causation. That is, it would be to simply insist, without argument, that mental causation does not violate any plausible exclusion principle – precisely what the proponent of the exclusion problem denies. In short, the proposed modification of the *unqualified exclusion principle* must be acceptable to reductive physicalists while remaining unavailable to dualists.

Turning to the final type of non-reductive physicalist response that is present in the literature: a few philosophers (like Lynne Rudder Baker (1993) and Carl Gillett (2003b) reject *completeness of physics* but seem to claim that this is consistent with

physicalism.[36] This seems like something that physicalists must avoid. As will become clear below, physicalists must endorse both *fundamental force completeness* (or a more general completeness claim of which it is a special case) and *causal sufficiency completeness* in order to distinguish their view from alternative views like dualism and emergentism.[37]

As noted above, each of these responses to the exclusion problem also undermines the standard formulation of the causal argument for physicalism. For, each denies one of the premises of the causal argument. However, there is a grain of truth in each of them. All of these responses are correct to claim that *causal sufficiency completeness* cannot be consistently combined with *causal efficacy* and a strong, unqualified version of the *exclusion principle*. Yet they stop short and miss the fact that *causal sufficiency completeness* and an *unqualified exclusion principle* can be consistently combined with, respectively, a *qualified exclusion principle* and *fundamental force completeness*. When each pair is combined with an appropriate *causal efficacy* claim, these two sets of theses make up a causal argument for physicalism, which I discuss in Chapter 3.

2.3.3. Generalization and the Exclusion Problem

Before developing my solution to the exclusion problem, I should discuss a non-reductionist response that does not explicitly deny any of its four theses. This is

---

[36] Baker claims that her position is not materialism and suggests rejecting what she calls the "metaphysical position" about causation. But she does not endorse dualism or epiphenomenalism, either.

[37] Gillett (2003b) notes that his version of physicalism, which includes "strongly emergent" properties, is incompatible with the completeness of fundamental physics (where a property instance is strongly emergent if and only if it is realized and partially non-causally determines the causal powers contributed by at least one of the fundamental properties that realize it). However, he asserts that it is compatible with physicalism and with what Fodor calls the "generality of physics." I doubt whether these claims are true, for Gillett's position requires the existence of "fundamental laws, ineliminably referring to strongly emergent properties, that modulate the causal contributions of the microphysical properties in ways consistent with, but not captured by, the laws only referring to microphysical entities" (2003b, 52 n.25). This seems to deny that fundamental force completeness is minimally true of the microphysical domain.

the so-called "generalization argument," which attempts to show that the exclusion problem generalizes to all special science properties and thus leads to the absurd conclusion that only fundamental, microphysical properties are causally efficacious. I think that the exclusion problem does generalize but that this is not itself an adequate response to the exclusion problem. However, showing how Kim's response to the generalization argument fails points up a tension in his position. This tension is a symptom of a general problem that I will exploit in the next section to pose a dilemma for proponents of the exclusion problem.

Many philosophers have attempted to disarm the exclusion problem by pointing out that it not only arises for mental properties but also applies to all special science properties that are apparently causally efficacious (e.g. Burge (1993), Baker (1993), Van Gulick (1992), Bontly (2002), Gillett and Rives (2001), Schröder (2002)). Given that Kim's argument only turns on the claim that mental properties supervene on but are not identical to restricted-sense physical properties and the fact that the same holds for all broadly-physical properties, the exclusion problem will arise for each of these areas. So it seems that if the reasoning behind the exclusion problem is sound, then only fundamental restricted-sense physical entities are causally efficacious. Further, if there turns on to be no fundamental physical level, but rather an endless series of levels, each supervening on a lower level, then there are *no* causally efficacious properties. That is, if there is no fundamental level and the exclusion problem is perfectly general, then causal efficacy "drains away into a bottomless pit" (Block 2003). Since it is allegedly absurd to think that only properties of fundamental physical entities are causally efficacious or that there are no causally efficacious properties, there must be something wrong with the line of reasoning that leads to the exclusion problem.

Kim has offered two responses to the generalization argument. First, he has argued directly against generalization. This direct argument proceeds by claiming that generalization is blocked by two alleged facts: (1) that micro-structural properties of a given object do not supervene on the properties of (and relations between) the object's parts, and (2) that a single micro-structural property can be decomposed at any level of compositional detail. If there is no bottom level, this results in an infinite sequence of identity claims. If there is a fundamental level, then this amounts to the claim that there are micro-structural properties only at the fundamental physical level (that a micro-structural property decomposed at any higher level of detail (e.g. the atomic level) is identical to some micro-structural property decomposed at the lowest microphysical level) (see Kim 2005, 68-9).

Kim's second response is that even if the exclusion problem generalizes the non-reductive physicalist still has work to do. The generalization argument takes the form of a *reductio ad absurdum*, so the non-reductive physicalist still needs to do two things: show exactly where the exclusion problem goes wrong and give a positive account of how the multiple simultaneous causally efficacious property instantiations are related to one another.

I think that this second response is a good one. This chapter attempts to accomplish the first of these tasks: show exactly how the exclusion problem goes wrong. The rest of the dissertation offers a positive account of how mental (and other special science) entities are related to restricted-sense physical entities and develops some of the implications of this account.

However, I think that Kim's first, direct response fails. This response begins with a distinction he draws between "levels" and "orders" of properties. Roughly, levels track the macro-micro distinction and are ordered by the part-whole relation; properties at different levels belong to different objects. By contrast, properties at

different orders belong to the same object. Kim apparently thinks that orders roughly track the distinction between the sciences; a single object can have different micro-structural properties of different orders, depending on the "level of detail" in which the object is decomposed.

I have doubts about whether the distinction between levels and orders is cogent and whether it has any bearing on whether the exclusion problem generalizes. However, setting these aside and assuming that levels and orders are independent of one another, we have two dimensions along which the exclusion problem might generalize: an inter-level (intra-order) problem and an intra-level (inter-order) problem.

Kim claims that micro-based or micro-structural properties – properties of being completely decomposable into (scientifically significant) proper parts that are propertied and related in a certain way – are crucial to blocking generalization along both dimensions. He believes that the inter-level exclusion problem – conflict between the properties of a whole and the properties of its parts – does not even arise. This is because of the alleged fact that supervenience is an *intra-level* relation, a relation between properties that all belong to the same objects and "hence are on the same level" (Kim 1998, 85-86). Hence, micro-based properties do not supervene on their constituents, the properties of entities at the micro-level. Thus the exclusion problem does not apply to them.

I think this response is a red herring; it depends on the technical details of an excessively narrow account of supervenience according to which a supervenient property and its base property must belong to the same object. This response does not work if we use the more general idea of dependence or determination which accounts of supervenience are supposed to capture. In fact, Kim himself at one time seemed to recognize this point. For his notion of mereological supervenience (and more

generally, supervenience for multiple domains) (1984, 96-971988, 113, 123ff.) captures the sense in which micro-structural properties of a macroscopic object are supervenient on the properties of the microscopic objects that are parts of that object. He apparently now thinks that this was merely loose talk, and that he should have said that the properties of an object supervene on *its* micro-structural properties. I see no reason to make this stipulation.

What about intra-level (inter-order) generalization – conflict between higher-order and lower-order properties of a single object? It seems that if the reasoning behind the exclusion problem is sound, then properties of the same macro-object at different orders will be in competition with one another (for example, my biological, molecular, and atomic properties). Here, causation does not drain away from the properties of macro-objects to those of micro-objects (i.e. from level to level). Rather, since an object at a given level on the macro/micro hierarchy can be decomposed in greater or lesser detail (i.e. can be decomposed into properties at different orders), we have the worry that micro-structural properties at one order will be preempted by the micro-structural properties at the lowest order on which the former supervene. Discussion of this possibility in his 1998 leads Kim to a position according to which higher-order properties are properties in name only (are merely property designators). This seems in effect to concede that this kind of generalization *does* occur. The only causally efficacious micro-structural properties are those that are decomposed at the level of fundamental physics.

However, Kim seems uncomfortable with this result and quickly tries to expand the domain of properties that count as physical (which seems to be a concession to the non-reductive intuitions present in the quotation from Kim 1998, 113-14). This suggests that Kim is ambivalent about whether or not the restricted-

sense physical domain includes any aggregate entities and their structural properties.[38] On the one hand, in order to save macrocausation and prevent the exclusion problem from generalizing, he needs to assume that the restricted-sense physical domain includes all of the properties of macroscopic objects which commonsense and the special sciences treat as causally efficacious. In other words, he must assume that complex, structural properties get in on the physical ground floor: they count as restricted-sense physical entities because they are "composed" of simpler restricted-sense physical properties and relations. This allows Kim to claim that: "Physicalism need not be, and should not be, identified with microphysicalism" (Kim 1998, 117).[39] But in that case, if macrocausation in general does not pose a problem, it is hard to see why mental causation does. Whatever operation (aggregation, realization, etc.) lets in structural and other macro-properties of macroscopic objects (and hence macrocausation) will threaten to let in mental properties of macroscopic objects (and hence mental causation) – bracketing the worry that realization is itself sufficient for reduction.

This prospect pushes Kim in the other direction. In order to make the exclusion problem as strong and persuasive as possible, Kim seems forced to claim that the restricted-sense physical domain does not include any aggregate entities or their structural properties.[40] However, without structural properties to appeal to, Kim is now forced to say that the exclusion problem is perfectly general. If there is no mental causation, then there is no macrocausation either. As Kim puts it:

---

[38] Kim seems to acknowledge this as a possible sticking point when he writes that Oppenheim and Putnam's hierarchical levels framework requires "that each level includes all mereological aggregates of entities at that level (that is, each level is closed under mereological summation). Thus, the bottom level of elementary particles, in this scheme, is in effect the universal domain that includes molecules, organisms, and the rest" (2005, 65 n.34). But in the main text, he does not appear to assume that this framework is operative nor notice the problems it raises for his position if it is.

[39] See also Kim 2005, 56.

[40] As, in effect, Trenton Merricks (2001) does, who holds that the exclusion problem raises problems for macrocausation in general.

> Assume that [the fundamental physical] level is causally closed; the [exclusion] argument, if it works, shows that mental causal relations give way to causal relations at this microlevel. And similarly for biological causation, chemical causation, geological causation, and the rest. So as far as the supervenience argument goes, the bottom level of fundamental particles (assuming that this is the only level that is causally closed) is always the reference physical domain. (2005, 66)

In the next section, I argue that the exclusion problem does not arise whether or not the restricted-sense physical domain includes aggregates and their structural properties. Before making that argument, I need to consider how to formulate the exclusion problem in light of the two types of completeness and the senses of "physical" discussed above.

2.4. The Exclusion Problem: A Second Pass and Solution

How do *fundamental force completeness* and *causal sufficiency completeness* and the two notions of physicality affect the formulation of the exclusion problem? First of all, it is clearly *causal sufficiency completeness*, not *fundamental force completeness*, that is relevant to the exclusion problem. After all, it is no part of non-reductive physicalism that there are fundamental mental forces. (*Fundamental force completeness* will play an important role in Chapter 3 in the part of the causal argument that rules out certain forms of dualism and emergentism.)

Since *causal efficacy* and *irreducibility* must be accepted by the non-reductive physicalist, we can determine what notions of physicality must appear in them and use these theses as fixed points to formulate *causal sufficiency completeness*, so that the three claims are inconsistent when conjoined with the *exclusion principle*. First, consider *causal efficacy*. While it is questionable whether mental events have restricted-sense physical effects (like collapsing a quantum wave function), it is uncontroversial that those at whom the exclusion problem is directed believe that mental events have broad-sense physical effects. This agrees with the literature on completeness. It is customary to single out the physical properties that are needed to

causally explain a "pre-theoretically given class of paradigmatic physical effects, such as stones falling, the matter in our arms moving, and so on" (Papineau 1993, 30). To be safe, we can take the set of effects to be the set of inclusive-sense physical effects.

What about *irreducibility* and the claim about physical sufficient causes that appears in *causal sufficiency completeness*? As noted above, non-reductive physicalists should accept that mental properties are inclusive-sense physical; the debate is whether or not they identical to restricted-sense properties. So, *irreducibility* has to be formulated in terms of the restricted-sense physical. Further, as discussed in the previous section, if the exclusion problem generalizes to at least some broad-sense physical properties, then we can think of it, quite generally, as attempting to demonstrate a conflict within the inclusive-sense physical domain between high-level and restricted-sense physical properties.[41] Versions of *causal sufficiency completeness* that appeal solely to the inclusive physical domain will not get any traction on the reductionists' project. This means that *causal sufficiency completeness* must be the claim that every inclusive-sense physical event has a restricted-sense physical cause. Putting all of this together, the more precise formulation of the exclusion problem is as follows:

> (CSC_{RS}) *Causal Sufficiency Completeness*$_{RS}$: For every inclusive-sense physical event and every time *t*, if that event has sufficient cause at *t,* then it has a restricted-sense physical sufficient cause at *t*.
>
> *Causal efficacy*: Mental events are causally sufficient for some inclusive-sense physical events.
>
> *Unqualified exclusion principle*: For every time *t,* (a) there cannot be two or more sufficient causes at *t* of an inclusive-sense physical event, except for cases of overdetermination, and (b) instances of mental causation are not, in general, cases of overdetermination.

[41] Even if the exclusion problem did not generalize to any broad-sense physical properties, this would not be simply because those properties were broad-sense physical.

> *Irreducibility*: Mental properties are distinct from (i.e. not identical to) restricted-sense physical properties.

As discussed above, (CSC_{RS}) is true only if the restricted-sense physical domain includes some aggregate entities and their properties. Again, instantiations of properties of aggregate objects result in causal powers that are do not result when only individual microphysical entities are present. For example, a 8 kg mass produces a gravitational force that no smaller mass does; a diamond is transparent and refracts light, but individual carbon atoms are not transparent and do not refract light (or at least not in the same way). These entities (and the causal powers they contribute) must be included in the complete physical domain if (CSC_{RS}) is to turn out true.[42]

Given this necessary condition for the truth of (CSC_{RS}), the following dilemma can be put to the proponents of the exclusion problem. Either the restricted-sense physical domain includes some aggregates and structural properties or it does not. If it does not include these entities and their properties, then (CSC_{RS}) is false. If it does include these entities and their properties, then (CSC_{RS}) may well be true. But in that case, I argue that the *unqualified exclusion principle* is false, even by reductionists' light. So, in either case the exclusion problem does not arise.

2.4.1. Assume that the restricted-sense physical domain does not include aggregate entities

Under the assumption that the restricted-sense physical does not include any aggregates or any of their properties, restricted-sense physical events are limited to changes in the properties of individual, fundamental restricted-sense physical entities. On this view, complicated structural entities are not restricted-sense physical entities. If one adopted this view, then the *unqualified exclusion principle* would apply to restricted-sense physical events. In cases where a restricted-sense physical event of

[42] It is perhaps somewhat misleading to continue to call this a restricted-sense physical domain, since it includes molecules, cells, organisms and their structural properties.

this kind (e.g. one that involves a change only in some property of an individual electron) causes an inclusive-sense physical event, it is plausible that there is only one restricted-sense physical sufficient cause. However, (CSC_{RS}) is plainly false on this understanding of restricted-sense physical events. As Kim himself notes, "The shattering of the glass was caused by the baseball and certainly not by the individual particles composing it" (2003, 167). (CSC_{RS}) is false even if we identify the baseball with a "complex structure of microparticles" as Kim does (see Kim 2005, 56). For, on the interpretation under consideration such "complex structures" are not part of the restricted-sense physical domain.

In making this claim, Kim is clearly committed to the existence of some macroscopic, aggregate physical objects. Recently, some philosophers have questioned this commitment and endorsed the view that only simple, fundamental particles (and perhaps living organisms) exist. One way to motivate such a view is to admit that the exclusion problem generalizes. According to this view, contra Kim, the "individual particles," arranged-baseball-wise, did cause the glass to shatter, and that the truth of this claim does not commit one to the existence of baseballs. To reach this conclusion one needs to argue that plural quantification and predication can be used to state all of the causal facts in the physical domain and that this plural quantification and predication does not commit one to the existence of aggregate entities. According to this perspective, (CSC_{RS}) is true of the maximally restricted-sense physical domain, and the exclusion principle rules out all alleged instances of macrocausation. In effect, this is to accept the purportedly absurd conclusion of the generalization argument; causation occurs only at the microphysical level (if it occurs anywhere).

I think that this conclusion *is* absurd, but as I indicated above, pointing this out is not a complete solution to the exclusion problem. However, I want to avoid the unfortunate bruiting of conflicting opinions regarding whether or not this is an absurd

consequence. I have three things to say in response. First, although a detailed discussion of plural quantification and predication would take me too far afield, I doubt whether plural quantification is ontologically innocent (i.e. whether it commits one to any entities in addition to the individuals quantified over in first-order logic) (see, e.g., Uzquiano 2004 for doubts about whether the logic of plurals is ontologically innocent).

Second, we should be more certain of the claim that events involving properties of macroscopic, aggregate objects are causally sufficient for other physical events than we are of the controversial metaphysical issues surrounding the ontological commitments of plural predication and quantification. After all, we acquire the concept of causation from paradigm macrophysical occurrences and perhaps from instances of mental causation (moving one's own limbs, etc.). It is plausible that any more refined concept of causation developed in philosophy or natural science will have to be continuous with and respect these paradigm cases.

Third, and most importantly, the proponent of the "no aggregate entities" view actually agrees that events involving changes in individual microphysical entities are not enough to provide a domain of which (CSC_{RS}) is true. In addition to these "atomic" events, we need specific relations between simple entities – relational properties picked out by such expressions as "arranged baseball-wise," which are not truly predicated of individual entities but only of some of them taken together. (After all, the same particles would not have broken the window if they were dispersed like a gas.) Given that this is so, I think that this kind of view is vulnerable to an argument of the same form that appears below in Section 2.4.2. That is, although this view is apparently not committed to the existence of complex aggregate objects or events (although, again this is debatable), the fact that they must admit that the *structure* of

the simple micro-entities is causally relevant means that their view is vulnerable to the same kind of argument I develop in the next section.

To anticipate, this line of argument runs as follows when applied to the "no aggregate entities" view. Assume that some arrangement or plurality of particles is causally sufficient for the breaking of a window (where talk of "arrangements" or "pluralities" need not commit one to complex, aggregate entities). In most cases, some slightly different *actual* arrangement of particles (e.g., one involving a couple fewer particles) is also causally sufficient for the window breaking. So, anyone who holds this view must also recognize that the *unqualified exclusion principle* is too strong. Any plausible exclusion principle must allow for cases of overdetermination other than those like firing squads (i.e. allow for cases in which there are two simultaneous sufficient causes of a given effect but in which these sufficient causes are not part of independent causal processes). I turn now to developing that argument in more detail for the case in which the reductive physicalist, like Kim, admits that some aggregate objects/events exist.

2.4.2. Assume that the restricted-sense physical domain includes some aggregates

Given a broadly physical event *B*, say, one's arm moving, what does one of the restricted-sense physical events that are causally sufficient for *B* look like? As noted above, it is not a simple microphysical event like an electron being excited to another energy level. Rather, it is an immensely complicated "aggregate" restricted-sense physical event – a number of microentities propertied and related in certain ways. Call such an event *A*, say, a collection of molecules in one's brain propertied and related in certain ways. Neuroscience suggests that other simultaneous events in the brain are also causally sufficient for one's arm moving. For instance, an event, *A'*, occurs that is exactly like *A* except that a single neuron fails to fire. In fact, there will apparently be countless such restricted-sense physical events, *A''*, *A'''*, etc., that occur and differ

only minutely from *A* but are still apparently causally sufficient for *B*. We are faced with a situation analogous to the "problem of the many" (see Unger 1980) – millions of sufficient causes where the proponents of the exclusion problem claim there can only be one, just as the problem of the many raises the prospect of millions of objects where there is intuitively only one. What should the friends of the exclusion problem say about *A*, *A'*, *A''*, etc.?[43]

One way to respond to this situation is to reject what Jonathan Schaffer (2003b) calls "individualism" with respect to overdetermining causes and adopt "collectivism." That is, one could claim that none of the individual *A*s is a sufficient cause of the arm movement but that only their mereological sum (or disjunction) is. While I think that Schaffer provides some strong arguments for individualism and against collectivism, it is clear that proponents of the exclusion problem cannot endorse collectivism. For, they *use* the assumption that individual apparent overdeterminers (the mental and physical events) are each sufficient causes as part of their argument against non-reductive physicalism's account of mental causation.

So, given individualism, the defenders of exclusion problem must claim that only one of *A*, *A'*, *A''*, etc. is a sufficient cause of *B*. How can they support this claim? It seems that they would need to provide a reason for choosing one of *A*, *A'*, *A''*, … as the sufficient cause. If one of these *A*s is the sufficient cause of *B* and the others are

[43] Perhaps one might claim that the minute differences between *A* and *A'* are not sufficient to make them distinct events. However, even using a fairly coarse causal criterion of event individuation, such minute differences will be sufficient for event distinctness. *A* and *A'* will have some different restricted-sense physical effects (e.g. they may have different effects on a sensitive charge reader), even if they have exactly the same broadly physical effects. One could insist on a very coarse-grained account of events according to which 'A' and 'B' refer to a single event if there is some token causal process which has a component that is accurately described by both 'A' and 'B'. In that case, all of the descriptions of the *A*s would pick out the same event, but so would any description of the mental event determined by them. (Assuming that there can be broadly physical and mental causal processes, and not merely microphysical ones.) One could consistently endorse an unqualified exclusion principle in this case, but it would not rule out mental causation (or macrocausation). My main point is that mental events fare no worse with respect to the exclusion problem than these complex restricted-sense physical events.

not, then there must be something that makes it the case that that restricted-sense physical event is the sufficient cause and the others are not. It seems that this kind of causal fact cannot be entirely arbitrary, but it is hard to see what sort of reason could be given.

One kind of response to the problem of the many denies the need for a principled reason in this context. It claims that although there is no reason for choosing one of the many over the others as the ordinary object, there is in fact only one ordinary object. The proponents of the *unqualified exclusion principle* might attempt to make the analogous claim: that there is no principle that determines which of the *A*s is the sufficient cause but that, nevertheless, only one *is* the sufficient cause. This could be done in several ways. First, they could claim that it is a brute metaphysical fact about causation that one of the *A*s is the sufficient cause of *B*, a position analogous to mereological "brutalism." Alternatively, they could claim that it is a brute semantic fact, which is in principle impossible for us to know, that, say, *A''*, satisfies "is a sufficient cause of *B*" and thus adopt an epistemicist solution. Finally, they could adopt a supervaluationist solution. On this approach they would admit that there is no fact of the matter about whether *A* or *A'* or *A''* or … lie in the extension of "sufficient cause," but they would claim that this is consistent with exactly one of the *A*s falling in that extension.[44]

[44] What about other approaches to the problem of the many? Clearly, nihilism about sufficient causes is not an option for the defenders of the exclusion problem nor is claiming that constitution is not identity. These approaches would be self-defeating. They may try the relative identity approach and claim that although *A* and *A'* are not the same restricted-sense physical event, they are the same sufficient cause. In addition to the standard objections to relative identity, this approach cannot simply say that *A* and *A'* are the same sufficient cause, for in some contexts their differences may be causally relevant (see note 43). Rather, it must say that they are the same sufficient cause of *B*, and thus make identity relative to very narrowly construed "sortals." Partial identity is not a plausible option either. While it may be that we count by "almost identity" when we talk about everyday things like cats and clouds, it is ad hoc to claim that we do so when talking about restricted-sense physical causes. This is related to a crucial difference between the problem of the many for ordinary objects and the problem that faces the defender of the exclusion problem. In the former one is trying to accommodate and defend a commonsense belief – for example, that there is one cat on the mat. In the latter, one is defending a

However, in the context of the exclusion problem one cannot turn to any of these solutions. For, the proponents of the exclusion problem *recognize* that they need to provide a reason for choosing one purported sufficient cause over another. *Causal sufficiency completeness* was supposed to serve as such a principle, as giving us a reason to choose the physical event over the purportedly distinct mental one as the sufficient cause.[45] But the friends of the exclusion problem obviously cannot turn to completeness to choose between the *A*s because they are all *within* the restricted-sense physical domain, which they claim is causally complete. Put another way, if one ultimately needs to appeal to arbitrary stipulation or brute facts in order to meet the challenge posed by restricted-sense physical causes like the *A*s, why bother with *completeness* as a premise in the exclusion problem at all? Why not just stipulate that there can be only one restricted-sense physical cause of a physical event at a given time and that *only* restricted-sense physical events can be causes of physical events? But this sort of stipulation would beg the question against non-reductive physicalism. The fact that the proponents of the exclusion problem cannot take this route prevents them from invoking any of the solutions in the last paragraph, each of which amounts to such a stipulation.

Reductive physicalists may claim that formulating things like this is tendentious in two ways. First, they might respond as follows: "What do you mean that completeness gives us a reason to 'choose' the physical over the mental cause? I claim that there is only one sufficient cause here; there are not two distinct property instantiations to choose between in the first place. Your *tu quoque* against my position

---

theoretical view to which the defender of the exclusion problem is driven, so one cannot appeal to some everyday folk concept or usage that must be preserved.

[45] "[The exclusion principle] itself is neutral with respect to mental-physical competition; it says either the mental cause or the physical cause must go, but does not favor either over the other. What makes the difference—what introduces an asymmetry into the situation—is [completeness]. It is the causal [completeness] of the physical world that excludes the mental cause, enabling the physical cause to prevail" (Kim 2005, 43).

simply begs the question."[46] This is a fair reaction. I have an *ad hominem* response that leads to a more substantive one. The *ad hominem* response is that Kim himself sets up the exclusion problem in terms of "choosing" the physical cause over the mental one (2003, 158). He does so for good reason. The reductionist is trying to argue that non-reductive physicalism is inconsistent. So he must assume for the sake of argument that there are two simultaneous sufficient causes, one physical and one mental and attempt to show that this leads to a conflict with the completeness of physics, giving us reason to conclude that the mental event is not a sufficient cause (unless it just *is* the physical event under another name).

Second, reductive physicalists may ask why they need to provide a principle to choose among the *A*s. Granted, they use *completeness* to rule out independent, *sui generis* mental events as causes, but why think that the same principle, or indeed *any* principle, is needed to solve the "problem of the many" involving restricted-sense physical causes? Why can't the reductive physicalist adopt one of the strategies proposed above for solving the problem of the many, and then, with this solution in hand, proceed to pose the exclusion problem (with the *completeness of physics* as a component) as a challenge to non-reductive physicalists?[47]

It may be true that the reductive physicalist need not rely on *completeness* to choose between the various *A*s, which would show that the discussion above is oversimplified. Perhaps it is also true that the reductive physicalist who is motivated by the exclusion problem need not provide any principle to choose between the *A*s and thus may adopt one of the strategies outlined above. But, even if this is the case, the reductive physicalist who takes this route must still claim that one of these "brute fact" strategies works in the restricted-sense physical domain but *not* in a more inclusive

[46] Thanks to Brian Weatherson for suggesting this line of response.
[47] Thanks to Andy Egan for raising this response and for general discussion about the dialectic at this point in my argument.

domain that includes mental properties. According to this move, some principle like *completeness* is still needed in the latter domain.

Reductive physicalists note that mental causation appears to involve a kind of overdetermination. They then use the claim that physics is causally complete to give precedence to the physical. I have argued that overdetermination is ubiquitous even within the restricted-sense physical domain, and, as will emerge below, I claim that the strategy that will accommodate restricted-sense physical overdetermination will *also* accommodate overdetermination between mental and restricted-sense physical causes. By contrast, the reductive physicalist claims that there is a "brute fact" strategy that accommodates overdetermination only in the restricted-sense physical domain, while a separate set of theses (the exclusion problem) applies in a domain that includes mental entities. This disjunctive treatment is unmotivated. The reductive physicalist has in effect been forced to make a distinction between two "exclusion problems" – one, the problem of the many that I have raised, which is claimed to arise only in the restricted-sense physical domain and which is resolved via some version of the "brute fact" approach – the other, which incorporates a completeness of physics claim, and which is used to rule out causation by unreduced mental events. But the reductive physicalist has given no reason why such a distinction should hold, other than that it leads to the result the reductive physicalist wants.

Thus, I think that this move by itself begs the question against non-reductive physicalists and dualists. The *exclusion principle*, and the apparent problem of which it is a part, should apply universally. Whatever plausibility this principle has derives from general considerations about causation and explanation – considerations that also apply in the restricted-sense physical domain. Why would simultaneous events compete as sufficient causes if one was broadly physical and the other restricted-sense physical but be compatible if both were restricted-sense physical? The exclusion

principle's plausibility derives from the general claim that there cannot be more than one independent causal chain or process leading to the same event (see Kim 1989, 243 n.15). Events like *A* and *A'* do not violate this claim. They are only minutely different from each other and thus there is some causal process of which they are both part that is sufficient for *B*.[48] Hence, they do not compete with one another for causal sufficiency. But then any broadly physical or mental event that is simultaneous with the *A*s *and that is part of the same causal process* will not violate this claim either and will not compete for causal sufficiency with one another or with the *A*s.

So, the reductionist needs to qualify the exclusion principle so that it allows for multiple simultaneous restricted-sense physical sufficient causes that are intimately related to one another. I claim that the intimate relation is that each must be part of overlapping token causal processes or part of the same token causal process that result(s) in the effect in question. An appropriate version of the exclusion principle must rule out only multiple simultaneous physical causes that are part of completely distinct token causal processes and thus are (in one good sense) independent of one another (again, setting "independent," firing-squad-like cases aside as different from mental causation):

> *Qualified exclusion principle*: For every time *t*, (a) there can two or more sufficient causes at *t* of a physical event only if they are part of overlapping token causal processes or part of the same token causal process leading to that event, except for cases of "independent" overdetermination, and (b) instances of mental causation are not, in general, cases of "independent" overdetermination.

However, using this exclusion principle does not rule out causation by mental or broadly physical properties. For, as I discuss in later chapters, there are token causal processes of which broadly physical events, including mental ones, and their

[48] Again, there will arguably be some causal pathways, e.g. those leading to restricted-sense physical effects, that *A* enters into but *A'* does not, which is why they are distinct events. See note 43.

restricted-sense physical realizers are both parts. Thus, I claim that a qualified exclusion principle that is acceptable on the reductionist's own terms will not rule out mental sufficient causes.

Importantly, this qualification is not available to the dualist. According to the dualist, mental properties are sui generis; they are not part of the physical world. This is why it is so hard to understand how they could causally interact with physical events. If dualism is to be a distinct position from physicalism, it must hold that mental causation is not a species of physical causation but a distinct kind of causal process, not explainable by the physical sciences. If there is to be more than a mere difference in degree between the mental and the physical, then it is hard to see how the causal processes bridging the mental/physical gap could be of the same kind as causal processes involving solely physical entities. For example, it is plausible that, according to the dualist, causal processes between mental and physical events are not spatiotemporally located (or at least not in the same way) as wholly physical causal processes are. Hence, mental and physical events cannot be part of overlapping token causal processes, and the dualist cannot avail herself of the *qualified exclusion principle* in order to make sense of mental causation. She cannot adopt the physicalist strategy of bringing the mental into the (inclusive-sense) physical fold.

The importance of formulating the exclusion principle so that it allows for overlapping token causal processes can be brought out by considering cases like the following. Consider a case where an invading horde brings ten battering rams to break down the castle gate, when any eight would be sufficient.[49] Here we apparently have a case in which there are forty-five ($_{10}C_8$) simultaneous overdetermining causes: all of the different ways of choosing (without replacement) exactly eight rams hitting from

[49] Assume that each of the battering rams strikes the gate at the same time. Thanks to Brian Weatherson for first suggesting this kind of case.

the ten that hit at the same time. Further, these events are not clearly part of completely independent causal processes (as the bullets are in a classic firing squad case). For any two of these events, one cannot identify two spatiotemporally disjoint, independent pathways that are each causally sufficient for the destruction of the gate. However, on the other hand, this kind of case does not appear to be a good model for mental causation, either. One cannot say that any of these events is determined by any of the others. No event is a spatiotemporal or logical part of or determines any other. Rather, these events merely spatiotemporally overlap one another.

These considerations pull in two directions. On the one hand, this is a case of overdetermination set up by conscious planning, and it would be strike us as improbable if it were widespread in the non-intentional natural world. So we might be inclined to say that it is an instance of "pernicious," independent overdetermination. This goes along with the idea that there appears to be no single token causal process of which any two of these events is a part.

On the other hand, cases like this one must be used to motivate the problem of the many presented above. If they are not used – if one only used cases where one apparently sufficient cause is a proper part of another – then one could simply appeal to the *minimal* sufficient cause to avoid overdetermination. So, if the problem of the many is to be effective against the proponents of the exclusion problem, then the exclusion principle must be qualified in such a way that these kinds of cases are allowed. That is, contra the first inclination, we apparently cannot assimilate this kind of case to firing squad cases.

So I think that the non-reductive physicalist should take the second option. That is, we need to not only allow cases in which the two simultaneous events are part of the same token causal process but also allow cases in which the simultaneous apparent sufficient causes are part of distinct token causal processes that

spatiotemporally overlap one another. The non-reductive physicalist will still think that this kind of case is not a good model for mental causation (because there is no determination relation, but only one of overlap, between the overdetermining events and the processes of which they are a part). But that does not pose a problem. There may be cases of "non-independent" overdetermination that differ in important ways from the mental/physical case. What they all share (and in this they differ from the pernicious, firing-squad-like cases) is that there is only a single set of overlapping (in some cases, completely overlapping) token causal processes that leads to the event in question.[50]

2.5. Concluding Remark

In this chapter, I have argued that the exclusion problem can be solved by paying careful attention to what is meant by saying that physics is complete. This is an improvement over existing response to the exclusion problem in that it provides the resources to develop a causal argument for physicalism that need not be reductive in form. In the next chapter, I develop this two-part causal argument.

[50] Note that claiming that simultaneous events do not compete for causal sufficiency only if they spatiotemporally overlap is too restrictive. Although this might work for events like the *A*s, it will not allow for certain cases of "quantitative overdetermination" (see note 35) which involve a single causal process (and thus no conspiracy or coincidence) but in which the sufficient causes are spatiotemporally disjoint (e.g. two halves of boulder that breaks a window). Thus, such a proposal is too restrictive.

CHAPTER 3

A CAUSAL ARGUMENT FOR PHYSICALISM AND THE DISTINCTION BETWEEN REDUCTIVE AND NON-REDUCTIVE PHYSICALISM

3.1. Avoiding Hempel's Dilemma

I shall introduce my two-part causal argument for a form of physicalism that need not be reductive by discussing, in broad outline, how the non-reductive physicalist should avoid Hempel's dilemma. First of all, one cannot determine *a priori* the domain of entities of which *causal sufficiency completeness* is minimally true. In the last chapter, I argued, following a passage from Kim, that the best available empirical evidence suggests that that domain must include at least some complex, aggregate restricted-sense physical entities. We cannot appeal merely to individual restricted-sense physical objects and their properties in order to find the domain of which *causal sufficiency completeness* is minimally true.

However, for all we know *a priori*, the smallest domain of aggregate physical entities of which *causal sufficiency completeness* is true may include aggregate entities that involve fundamental *mental* forces. Second, for all we know *a priori*, mental properties may be needed in the smallest domain of which *causal sufficiency completeness* is true. Either of these options threatens to make physicalism a doctrine without much content. Again, physicalism that is defined in an "open-ended" way, by reference to future or ideal physical theory or without making any restrictions on the kinds of forces that can be involved in aggregates, gets stuck on the "vacuity/triviality" horn of Hempel's dilemma.

In a recent paper, Karen Bennett (forthcoming) has suggested that if the second claim is true, if *causal sufficiency completeness* is minimally true of the inclusive-sense physical domain, then this may itself provide a non-reductive physicalist

response to the exclusion problem.[51] After all, the non-reductive physicalist grants that mental properties are inclusive-sense physical, and "the non-reductive physicalist can perfectly well claim that some events have causes that are physical in the weaker, [inclusive] sense, but not in the stronger, [restricted] sense. That is, there would be room to claim that some events only have mental, or otherwise higher-level physical causes. Completeness would not guarantee that the effects of mental causes always have another physical cause lurking nearby" (Bennett forthcoming). If this were so, then the exclusion problem would not arise because it would equivocate on the sense of 'physical.' If *causal sufficiency completeness* is the claim that every inclusive-sense physical event has an *inclusive-sense* physical cause, then this is compatible with the *causal efficacy* of the mental, the *unqualified exclusion principle*, and *irreducibility* (which, again, is only the claim that the mental is not *restricted-sense* physical).

Bennett suggests that this is not a complete solution to the exclusion problem because a version of the problem simply reappears within the inclusive-sense physical domain.[52] However, I think that this proposed solution is unsatisfactory for another reason. This solution to the exclusion problem makes the corresponding causal argument for physicalism useless. If unreduced mental properties must already appear within the minimal causally complete domain, there is no need to appeal to *causal efficacy* or the *exclusion principle* to conclude that they are physical. But this means that the kind of physicalism that results from the equivocation solution to the exclusion problem itself gets stuck on the vacuity horn. We would have effectively

[51] I made a similar claim, independently, in a paper read at the 2004 meeting of the Creighton Club.
[52] But, she claims that "it is not quite the *same* version. And whatever happens, all apparent competition has been reduced to competition among *physical* events and properties" (Bennett forthcoming). Another way to put my point in this paragraph is that it is, in fact, the same problem. Without a constraint on the fundamental nature of the inclusive-sense physical, the non-reductive physicalist has merely made a terminological change.

defined 'physical' as 'whatever is needed to give a complete causal account of the world,' and this would not rule out fundamental vital or mental forces from appearing in that account (as strong emergentism claims).[53]

One standard strategy of avoiding the vacuity horn is to exclude mental properties from the physical domain by definition – the so-called "via negativa." This strategy may allow reductive physicalists to avoid Hempel's dilemma, but I think that non-reductive physicalists should be wary of it. For, it leads to the apparently inconsistent formulations of their position discussed in Chapter 2, according to which mental properties are *non-physical* and excluded from causal efficacy by the physical. However, I think that non-reductive physicalists can endorse a broadly similar strategy – one that has been often conflated with the "via negativa." Non-reductive physicalists should avoid triviality or the unacceptable extension of future physical theory by noting that maximally restricted-sense physical entities (in particular, those that are non-mental and non-vital) provide for a complete set of fundamental forces. The seed of this idea may be found in Papineau's empirical claim that current physics will not need to be supplemented by mental or vital "categories" in the future (1993, 31). Although Papineau does not mention this, note that the term "categories" here does not refer to the sufficient causes (i.e. events) of physical effects themselves but to the "building blocks" of those causes – things like "energy, field[s], and spacetime structure" (ibid.). That is, this empirical claim is no longer about sufficient causes but about fundamental forces (if forces are identified with fields) and other constituents of events. If *fundamental force completeness* is true, future physical theory cannot be

[53] Cf. Papineau: "Suppose we simply *define* 'physics' as the science of whatever categories are needed to give full explanations for all physical entities. ... [T]here is no difficulty about how we know that it is complete, for we have simply defined it so as to be complete. ... In itself, the above definition of physics leaves it open that psychological categories may turn out to be needed as an essential part of physics" (1993, 29-30).

extended in just any way and remain compatible with physicalism. For example, if dark matter turned out to be conscious, then physicalism would be false.

Suppose, as Bennett suggests, that we had empirical evidence or other reason to think that mental entities were included in the minimal physical domain of which *causal sufficiency completeness* is true. Then, the causal exclusion problem will not reappear in that inclusive-sense physical domain, if, as I argued in Chapter 2, only a *qualified exclusion principle* is plausible in that domain. However, that exclusion principle can be used, together with *causal sufficiency completeness* and the *causal efficacy* of the mental, in an argument for the claim that mental causes are not independent of the other physical causes in that domain. This is not yet physicalism, for we have not established that the mental is not fundamental. However, it is a kind of causal monism. A second argument involving *fundamental force completeness*, *causal efficacy* and an *unqualified exclusion principle* can then be used to complete the argument for physicalism – to argue for the claim that the mental is determined in an appropriate way by the restricted-sense physical.

On the other hand, suppose that we had empirical evidence that *causal sufficiency completeness* was minimally true of a physical domain that included complex physical aggregates and *some* of their properties, but not mental properties. Assuming that the mental is causally efficacious, the very same reasoning sketched in the preceding paragraph could be used to argue that the mental is not independent of the physical causes in the domain of which *causal sufficiency completeness* is true. We can then appeal to the same second argument mentioned above, involving *fundamental force completeness*, to ensure that it is the restricted-sense physical which grounds or determines the mental.

In the next section I fill in the details of this two-part response to Hempel's dilemma – showing how my solution to the exclusion problem provides the resources

to develop a causal argument for physicalism that need not be reductive. In Section 3.2.1, I discuss the first part of this causal argument, which guarantees the unity and exhaustiveness of the physical world. In Section 3.2.2, I discuss the second part, which ensures that physicalism is a "fundamentalist" doctrine – that mental forces are not fundamental forces, not basic ontological ingredients in the world.

3.2. A Two-Part Causal Argument for Physicalism

3.2.1. Causal Monism

As mentioned above, in Chapter 2, I argued that *causal sufficiency completeness* is minimally true of a restricted-sense physical domain that includes some aggregate entities and their properties but that the *unqualified exclusion principle* is not tenable there. Rather, in order to be consistent, the reductive physicalist must endorse a weaker, *qualified exclusion principle*. Putting these two premises together with the thesis that mental events are causally efficacious results in the following argument:

> *Causal sufficiency completeness*: For every inclusive-sense physical event and every time *t,* if that event has a sufficient cause at *t,* then it has a restricted-sense physical sufficient cause at *t* (where the set of restricted-sense physical entities includes some aggregate physical objects and their properties).
>
> *Causal efficacy*: Mental events are causally sufficient for some inclusive-sense physical events.
>
> *Qualified exclusion principle*: For every time *t*, there can two or more sufficient causes at *t* of an inclusive-sense physical event only if they are part of overlapping token causal processes or of a single token causal process leading to that event (except for cases of "independent" overdetermination, which can be set aside).
>
> Therefore, *causal monism*: Mental events (and their constitutive properties) are part of a token process that overlaps a token process involving restricted-sense physical events, and are thus not independent of, those restricted-sense physical events (and their constitutive properties).

Note that this conclusion does not entail that *irreducibility* is false; one can maintain that mental events are not independent of complex, aggregate restricted-sense physical events while at the same time claim that they are not identical to (simple or aggregate) restricted-sense physical events. However, this argument is not sufficient to establish physicalism.

First, although *causal monism* establishes that mental and restricted-sense physical events are not independent of one another, it does not establish whether it is mental properties that are determined by (dependent on) restricted-sense physical properties, or vice versa. *Causal monism* is compatible with restricted-sense physical events being determined by mental events, and not vice versa, a claim which obviously no physicalist can endorse. Further, it does not establish whether the mental and restricted-sense physical events merely overlap, e.g. are each partially determined by some third kind of event, as neutral monists might claim. Finally, cases of strong emergence, which involve certain configurations of restricted-sense physical entities giving rise to novel fundamental mental forces, also arguably are not ruled out by this argument. At least according to the view that natural laws are metaphysically necessary, emergent properties will be necessitated by their bases. In any event, such cases do not involve two independent causal processes that happen to result in the same effect, which characterizes the "pernicious" kind of independent overdetermination.[54,55]

[54] But don't emergentists deny *causal sufficiency completeness*? I think they need only deny it when the restricted-sense physical domain is interpreted as excluding aggregates or "configurations" of restricted-sense physical entities, i.e., when *causal sufficiency completeness* is formulated with "maximally restricted-sense physical" in the consequent.

[55] Also, *causal monism* cannot rule out epiphenomenalism and non-interactionist parallelism (or pre-established harmony views) because such views deny *causal efficacy*. However, no causal argument for physicalism rules out such views without some additional assumptions and argumentation (see, e.g. Lewis 1966).

In effect, this argument, taken by itself, avoids the falsity horn of Hempel's dilemma but gets stuck on the "vacuity/triviality" horn. We would have effectively defined 'physical' as 'whatever is needed to give a complete causal account of the world,' and this would not rule out fundamental vital or mental properties from appearing in that account. Without any further claim about the nature of the mental and its relation to the restricted-sense physical, the resulting "physicalism" about the mind would be merely empty rhetoric. In other words, physicalism threatens to become a trivial claim. One way of thinking of this is that although *causal sufficiency completeness* entails *fundamental force completeness* when both concern the same physical domain (as I argued in the previous chapter), *causal sufficiency completeness* construed as a claim about aggregates does not entail *fundamental force completeness* concerning the domain of simple, fundamental restricted-sense physical entities. The former is a weaker claim because it involves a broader class of entities.

I think that the way to respond to this is not to fuss with the sense of 'physical' involved, as some have done (see, e.g. Papineau 1993, 30-31). Rather, one should recognize that the completeness of physics has another part: *fundamental force completeness* (or the more general claim of *aggregational completeness*, of which *fundamental force completeness* is a special case). As I argue in the next section, the fact that *fundamental force completeness* is minimally true of a physical domain that does not include vital or mental entities is what prevents physicalism about the mind from becoming a trivial claim. That is, we need an additional argument involving *fundamental force completeness* and a version of the *unqualified exclusion principle* to guarantee that the resulting formulation of physicalism is substantive and continuous with materialist doctrines, and thus avoid the vacuity horn of Hempel's dilemma.

However, *causal monism* goes some way toward ruling out the views according to which empirical reality is divided into causally isolated domains. The

conclusion that there is only a single causal process (or overlapping causal processes) for any given inclusive-sense physical event rules out views according to which there are many causally isolated layers in the inclusive-sense physical domain: a world in which causally isolated levels run in harmonious lockstep with one another. It thus establishes a version of monism, but leaves open the nature of the "one kind of stuff" which makes up everything.

It also arguably rules out occasionalist forms of Cartesian dualism. Presumably, occasionalist dualists accept *causal efficacy*, if it is interpreted as the claim that mental events need only be occasional causes of inclusive-sense physical events. But such events would not be part of the same causal process as any restricted-sense physical events, for the mental events would not be located in space-time with restricted-sense physical events. (This assumes that some spatiotemporal relation between *A* and *B* is a necessary condition for *A* and *B* being part of the same token causal process or overlapping causal processes.)

3.2.2. Reducibility of Forces

Return to the first horn of the dilemma I posed in the previous chapter for the proponent of the exclusion problem. On this horn, the restricted-sense physical domain is assumed to not contain any aggregate physical entities or their properties; hence, causal sufficiency completeness is not true of this maximally restricted-sense physical domain. However, there is no reason to think that *fundamental force completeness* will not be minimally true of the restricted-sense physical domain interpreted in this way. Indeed, empirical evidence suggests that *fundamental force completeness* is minimally true of the restricted-sense physical domain.

David Papineau has argued that the past two hundred or so years of empirical science provides evidence for the claim that the conservation of energy holds universally (see Papineau 2001 and the appendix to his 2002b for a development of

this point). Now, the conservation of energy by itself does not establish *fundamental force completeness*. For any non-restricted-sense physical forces could be fundamental and compatible with the conservation of energy by being governed by deterministic laws and conservative (Papineau 2001, Section 8.iv). However, Papineau suggests that two lines of thought have led scientists to reject the existence of such forces. First, there is what Papineau calls an "abstract," inductive argument to the effect that since all apparently special forces have been shown to reduce to a small number of physical forces that conserve energy, we should expect mental forces to be reducible as well. I think that this argument is rather weak. After all, based on the same evidence one could offer the alternative inductive argument that, since all fundamental forces discovered so far have been conservative, we should expect that any fundamental mental forces would be conservative as well (cf. Papineau 2001, Section 9.iii). I think that the second line of argument is more important and more convincing. This argument appeals directly to the physiological evidence that there are no anomalous processes, phenomenon, or accelerations within living and sentient things that are not present in other matter (which would require novel fundamental vital or mental forces).

This physiological evidence supports the claim that there are no fundamental forces present in organic or sentient entities that are not present in inorganic entities. I think that this amounts to a strong completeness claim with respect to fundamental forces.[56] Only fundamental restricted-sense physical forces (and forces derivative from these) can affect physical entities; there cannot be causal interactions between fundamental restricted-sense physical forces and fundamental non-restricted-sense physical forces (if there were any).

[56] Where strong completeness for forces amounts to the claim that there can only be *restricted-sense physical* fundamental forces involved in the production of any inclusive-sense physical event. See note 25 in Chapter 2.

If this is right, then an *unqualified exclusion principle* is implicit in Papineau's historically based argument for physicalism. It is required to rule out any purported non-restricted sense physical forces that are independent of (not determined by) fundamental restricted-sense physical forces and combine with them to produce the same effects that the fundamental restricted-sense physical forces would have produced alone. But such an exclusion principle, in concert with *fundamental force completeness*, poses no threat to mental causes since the non-reductive physicalist claims that the mental and other broadly physical forces (like friction) are determined by fundamental restricted-sense physical forces.

This gives us the other part of an argument for physicalism (that need not be reductive) – an argument that rules out fundamental vital or mental forces. This can be thought of as an argument for the reduction of all forces that impinge on the inclusive-sense physical domain to fundamental restricted-sense physical forces.

> (FFC_{RS}) *Fundamental force completeness*$_{RS}$: Fundamental restricted-sense physical forces are sufficient for the causal processes leading to every inclusive-sense physical effect (where the restricted-sense physical domain contains no aggregate entities).
>
> *Force efficacy*: Mental entities engender forces that are relevant to inclusive-sense physical effects; they enter into causal processes that have inclusive-sense physical outcomes.
>
> *Unqualified exclusion principle*$_{FF}$: If some fundamental forces are sufficient for the causal processes leading to every inclusive-sense physical effect, then no other wholly distinct forces are part of (constitutive of) those causal processes.
>
> Therefore, *reducibility of forces*: Mental forces that impinge on the inclusive-sense physical domain are reducible to fundamental restricted-sense physical forces.

However, reducibility of forces is compatible with the claim that causal powers contributed uniquely by mental events (indeed, those contributed uniquely by any broad-sense physical events) are necessary components of a complete causal domain

in the sense given by *causal sufficiency completeness*. This means that this conclusion is compatible with irreducibility. Mental events and properties will be distinct from restricted-sense physical events and properties, assuming that a broadly causal criterion of event and property individuation is adopted, since mental and restricted-sense physical events and properties will have different causal profiles (contribute different causal powers).

Like the argument for causal monism, this argument by itself is not sufficient to establish physicalism. For, it is compatible with overdetermination by causal processes that ultimately involve only fundamental restricted-sense physical forces, and views according to which the inclusive-sense physical domain is divided up into causally isolated layers. It is also compatible with the existence of non-physical forces that are isolated from the inclusive-sense physical domain. That is, it is compatible with psycho-physical parallelism, views according to which there is a pre-established harmony between physical and non-physical events. In general, it does not rule out a bifurcated empirical reality in which causally isolated domains run in lockstep with one another. These views are ruled out by causal monism.

Thus, these two arguments thus work in tandem. Reducibility of forces ensures that calling mental entities 'inclusive-sense physical' is substantive and not merely a terminological recommendation. It rules out the kind of force emergentism and mental monism that is compatible with causal monism. Causal monism, on the other hand, sets limits on the kinds of causal relations that obtain in the world. There is no ontological distinction between properties of the kind that would underwrite dualism (and a corresponding methodological split between physical/natural sciences and human sciences). The empirical world is a unified whole, with one type of causation.

It is important to note that the only evidence we have for physicalism is the actual success of physical science in causally explaining physical occurrences. For this widely accepted fact to be plausible, we must recognize that the successes we have had in causally explaining the vast majority of physical phenomena, like the development and movement of animals and the products of chemical reactions, have not come from fundamental physics. (For example, there is no reductive account of life in terms of restricted-sense physical properties.) In order for such successes to be evidence for physicalism, we need to invoke a relatively broad notion of physical science, that is, any science which proceeds under the assumption of *fundamental force completeness*$_{RS}$ – that all fundamental forces are restricted-sense physical. In this sense, contemporary biology, chemistry, physiology, and psychology are all physical sciences. Thus, the empirical support for *fundamental force completeness*$_{RS}$ (and not anything to do with *causal sufficiency completeness*) is what ensures that physicalism is a genuine alternative to strong emergentism and dualism.

3.3. Relaxing the Assumption that there is a Fundamental Level: Aggregational Completeness

The two-part argument developed so far makes the question of whether physicalism is true turn on the question of whether there is a fundamental physical level and perhaps on whether there are fundamental vital or mental forces.[57] But this does not seem right; whether or not physicalism is true is plausibly independent of these issues. (Of course, it may be harder to state physicalism precisely if there is no bottom level (cf. Schaffer 2003a and Montero 2006).) The more general issue (of which the question of whether there are any fundamental vital or mental forces in a world with a bottom level is a special case) is whether any mental or vital entities

[57] But see Papineau (2001; 2002b, n.7, n.8) for discussion of the question of whether appealing to forces is merely an expedient assumption.

appear below a certain level of complexity. The question of whether there are any fundamental mental forces is a special instance of the question of whether mental entities are to be found at any arbitrary level of decomposition. Both of these questions are concerned with whether mentality is a deep feature of the world on par with restricted-sense physical features like mass and charge, or, if it is rather a feature which appears only in certain very complicated systems and which is asymmetrically dependent on more basic physical properties.

If physicalism is true, then all entities that have physical effects are determined by restricted-sense physical entities such that any further determination of these entities is also by restricted-sense physical entities (cf. Montero (2006, 178)). In other words, if physicalism is true, then one can find some level below which all determination is by restricted-sense physical entities. Call any (possibly one-way infinite) sequence of all of the entities that compose the constitutive objects and properties of a given event an "aggregation base." To obtain an aggregation base, we pick any level of complexity and then decompose the event into finer and finer levels of detail. If there is no fundamental level, no level at which we reach simples and their properties, then every event will have an infinite number of infinite aggregation bases (one for each level at which we choose to start the decomposition). The more general completeness claim of which *fundamental force completeness*$_{RS}$ is a special case is that every event has some aggregation base which is completely non-mental. At some point in the decomposition we reach a level below which there are no mental entities (no entities with mental properties). Roughly, the more general claim captured by this kind of completeness is that there is not mentality "all the way down."

> *Aggregational completeness*$_{RS}$: For every event that has an inclusive-sense physical effect, restricted-sense physical entities are an aggregation base of that event.

Since *fundamental forces completeness* is a special case of *aggregational completeness*, the former entails the latter, but not vice versa (assuming that the same senses of 'physical' are involved in each). Assume *fundamental force completeness* is true, then the entities from the fundamental level are restricted-sense physical and form a (trivial, one-level) aggregation base for the constitutive object of any given event that has an inclusive-sense physical effect. Thus, *aggregational completeness* is true. Clearly, *aggregational completeness* does not entail *fundamental force completeness*. If there are no fundamental physical entities or forces (e.g. if matter is infinitely decomposable), then *aggregational completeness* might be true but *fundamental force completeness* will be false.

Note also that *aggregational completeness* and *causally sufficiency completeness* are not equivalent (assuming that the same senses of 'physical' are involved in each). I take it that part of what is suggested by the long passage from Kim quoted in Chapter 2 is that *aggregational completeness*$_{RS}$ does not entail *causal sufficiency completeness*$_{RS}$. Even if the maximally restricted-sense physical domain exhausts all of the bases of aggregation, it need not include a sufficient cause of every physical effect. For, the causal antecedents of some physical effects involve magnitudes and properties that are not part of the microphysical domain (e.g. a quantity of mass had only by macrophysical objects, a chemical property possessed only by organic macro-molecules).

What reason is there to think that *aggregational completeness* is true of the maximally restricted-sense physical domain? In short, I think that the empirical physiological evidence, mentioned above, for the claim that the conservation of energy holds universally supports *aggregational completeness*$_{RS}$ in addition to *fundamental force completeness*$_{RS}$. There is no reason to think that there are entities in the

aggregation bases of mental phenomena that are not in the aggregation bases of non-mental phenomena.

Importantly, the evidence we have that physics (in the restricted-sense) provides a complete catalog of the bases of aggregation does not come solely from the work of particle physicists and cosmologists. This evidence comes from special sciences like physiology which have provided a physically acceptable account of previously suspect biological functions (like reproduction, respiration, protein synthesis, etc.) and have shown that energy is conserved in the interactions between living things and their environment. Indeed, ongoing work in cognitive psychology and neuroscience is relevant to establishing that seemingly non-physical phenomena, like perception, emotion, and thought, are ultimately grounded only in the same entities that make up non-sentient things.

This evidence supports a more general argument that plays the same role as the argument for *reducibility of forces* – an argument for the claim that the mental is ultimately aggregated solely out of restricted-sense physical entities:

> *Aggregational completeness*$_{RS}$: For every event that has an inclusive-sense physical effect, restricted-sense physical entities are an aggregation base of that event.
>
> *Unqualified exclusion principle*$_{AB}$: If some entities are, and hence are metaphysically sufficient for, an aggregation base of event X, then no other wholly distinct entities are part of that aggregation base of X or constitutive of X in some other way.
>
> *Causal efficacy*: Mental events are causally sufficient for inclusive-sense physical effects.
>
> Therefore, every mental event that is causally sufficient for an inclusive-sense physical effect is aggregated *solely* of restricted-sense physical entities.

Assuming that aggregation preserves (inclusive-sense) physicality (which I discuss, with respect to properties, in the next chapter, see note 22), this argument

ensures that calling mental entities 'high-level physical' is substantive and not merely a terminological recommendation. It rules out the kinds of strong emergentism and mental or neutral monism that are compatible with the argument for *causal monism*. However, this conclusion is compatible with the claim that the causal powers contributed by mental properties (indeed, those contributed by any high-level physical properties) are necessary parts of a complete causal domain in the sense given by *causal sufficiency completeness*. That is, this conclusion is compatible with *irreducibility* of mental properties. A given mental property will remain distinct from (i.e. not identical to) any restricted-sense physical property, as long as it is associated with a distinct set of causal powers – that is, as long as mental properties carve up the space of all the causal powers, both fundamental and derivative, differently than restricted-sense physical properties do.

While the scientific evidence supports the view that all bases of aggregation belong to a relatively small set of restricted-sense physical entities, there is no empirical evidence that *casual sufficiency completeness* and the *unqualified exclusion principle* are both true of the restricted-sense physical domain. Thus, there is no reason to think that all causally efficacious properties are identical to maximally restricted-sense physical properties. In fact, there are empirical reasons to think that these claims are not true of this domain. For, as mentioned above, the only evidence we have for physicalism is the actual success of the physical sciences in causally accounting for physical occurrences. We must recognize that the successes we have had in causally explaining the *vast majority* of physical phenomena *have not* come from fundamental physics. In order for such successes to be evidence for physicalism we need to invoke a relatively broad notion of physical science, that is, any science which is compatible with *aggregational completeness*$_{RS}$. In this sense, contemporary biology, chemistry, and psychology are all physical sciences (i.e., physical in the

inclusive sense). Thus, the empirical support for *aggregational completeness*$_{RS}$ (and not anything to do with *causal sufficiency completeness*) is what ensures that physicalism is a genuine alternative to strong emergentism and dualism.

Non-reductive physicalists who insist that the exclusion problem generalizes are right to claim that it leads to a scientifically implausible conclusion, one that is inconsistent with our best explanatory practices. I have tried to show that paying close attention to the different ways in which "physics is complete" (and the empirical evidence for each of these claims) not only shows exactly how the exclusion problem goes wrong but also leads us to a powerful causal argument for physicalism about the mental. Indeed, the fact that these arguments generalize can now be seen as a virtue, since they can be used to show that other seemingly non-physical, causally efficacious properties, like some biological and social ones, are inclusive-sense physical. And they are equally effective at showing the error in causal exclusionary reasoning that leads to the conclusion that biological processes (like natural selection) or social phenomena (like monetary exchanges) must be either epiphenomenal or identical to lower-level physical phenomena.

3.4. Explanatory Exclusion

However, there is good reason to think even the conjunction of *reducibility of forces* (or the conclusion of the argument involving *aggregational completeness*), *causal monism* and *irreducibility* (of broadly physical and mental properties) is not a complete answer to the reductionist challenge. Kim (1989) originally put the exclusion problem in terms of causal *explanatory* exclusion and completeness.[58] In that paper, Kim claims that if one cause is dependent on or derivative from another,

[58] In his argument for the identity theory, David Lewis also appeals to an explanatory version of the completeness of physics, which he calls the "explanatory adequacy of physics," the thesis that "there is some unified body of scientific theories, of the sort we now accept, which together provide a true and exhaustive account of all physical phenomena" (1966, 23). For an argument against explanatory exclusion, see (Marras 1998).

then the latter gains "explanatory or causal dominance over the [former]" (Kim 1989, 246). Further, he assumes that the restricted-sense physical causal account is always "deeper and more theoretical and systematic" (ibid., 251) and "more detailed, more revealing, and theoretically more fecund" (ibid., 249). One could still accept these claims while endorsing the arguments for *reducibility of forces* and *causal monism*. There are thus more reductive assumptions present in Kim's original exclusionary worries than are explicit in his recent (2005) reformulation of them.

Even if mental entities and their restricted-sense physical realizers do not compete for causal sufficiency, they may still compete with respect to causation – for being *a* or *the* cause of a given effect. That is, they may compete with a sense of causation that is sensitive to explanatory considerations.

Put another way, someone might respond to the argument thus far as follows: The reason that mental and some restricted-sense physical events do not compete for causal sufficiency is that only maximally restricted-sense physical properties enter into causal processes. There are no natural, mental (and other broadly physical) events and properties. Rather, there are only abstract, mental descriptions of basic, microphysical properties and causal processes that we happen to find convenient.[59]

In the next chapters, I argue that the right account of realization provides the means to show why this kind of response fails. To anticipate, sometimes the broadly physical causal account is deeper and more fecund because it isolates the relevant causal process (or mechanism) from a mass of irrelevant microphysical detail. Further, the broadly physical account may be more theoretical and systematic, since, as I argue, there will be no natural restricted-sense physical causal process (mechanism) that is unique to a given broadly physical effect.

[59] Whether or not this is true is the crux of the debates about the "autonomy" of special science explanations.

3.5. The Distinction Between Reductive and Non-reductive Physicalism

The additional problem posed by explanatory exclusion suggests a way of drawing the distinction between reductive and non-reductive physicalism. All physicalists are committed to the hypothesis that all of the entities in the inclusive-sense physical domain (chemical, social, biological, astronomical, geological, meteorological, mental, etc.) are ontologically unified – that they do not belong to distinct ontological domains. However, I think that reductive and non-reductive physicalists interpret this hypothesis in very different ways. According to reductive physicalists, all natural properties are restricted-sense physical. There are no natural special science properties (and hence no special science causal processes) but merely special science property designators – different ways of referring to restricted-sense physical properties. As domains of natural properties, the restricted-sense and inclusive-sense physical domains are held to be coextensive. The non-reductive physicalist denies this. She holds that there are natural physical properties that are not restricted-sense physical. The set of natural inclusive-sense physical properties cannot be identified with the set of natural restricted-sense physical properties.

I have argued that an adequate non-reductive solution to Hempel's dilemma and the exclusion problem must distinguish between different ways in which the domain of physical properties is causally complete: *fundamental force completeness* (or its generalization, *aggregational completeness*) and *causal sufficiency completeness*. I think that this distinction results in an ambiguity in the notion of a *natural property*. Natural properties are supposed to those that, *inter alia*, "characterise things completely and without redundancy" (Lewis 1986, 60). However, this could be interpreted as a claim about *fundamental force (aggregational) completeness* or *causal sufficiency completeness*. I use this ambiguity to articulate the core debate between reductive and non-reductive physicalists. Consistent reductive

physicalists (those that are not tempted by non-reductionist intuitions) must claim that there is a single set of properties (the restricted-sense physical ones) that is the smallest set of which both *fundamental force (aggregational) completeness* and *causal sufficiency completeness* are true. According to the reductive physicalist, the restricted-sense physical properties are the only natural properties. By contrast, non-reductive physicalists maintain that mental and other broadly physical properties are natural in that they are needed in the minimal set of properties of which *causal sufficiency completeness* is true (along with fulfilling the other roles of natural properties).

Some of Lewis's and Kim's comments about broadly physical properties like *pain* and *heat* suggest that differences of opinion regarding natural properties might be relevant to the distinction between reductive and non-reductive physicalism. For example, Lewis claims that the property of *being in pain* "cannot occupy ... any ... causal role because it is excessively disjunctive, and therefore no events are essentially havings of it. To admit it as causally efficacious would lead to absurd double-counting of causes" (1994, 307). Lewis also claims that the property (of space-time regions) *containing rapidly moving particles* is "a fairly natural, intrinsic property" while the property *containing whatever phenomenon occupies the heat-role* "is highly disjunctive and extrinsic." So, Lewis claims, events involving the former are genuine (natural) and causally efficacious, while events involving the latter are "too unnatural" and thus "inefficacious in the sense that [they cannot] figure in the conditions of occurrence of the events that cause things" (1983, 44-5).

Kim uses a similar line of reasoning to reach the same conclusion: "Given that mental kinds are realized by diverse physical causal kinds, therefore, it follows that mental kinds are not causal kinds, and hence are disqualified as proper scientific kinds. Each mental kind is sundered into as many kinds as there are physical realization

bases for it, and psychology as a science with disciplinary unity turns out to be an impossible project" (1992, 327). According to Kim, the heterogeneous and disjunctive nature of multiply realized mental kinds means that they are unnatural, unscientific and not causal.

So I am suggesting that the distinctively reductionist feature of Lewis's and Kim's philosophy of mind is not the (species-restricted) type-identity claims that they make about broadly physical properties and events, but rather the claim that broadly physical properties and events are not natural. This way of drawing the distinction holds that the core dispute between non-reductive and reductive physicalism is about which properties are natural and, ultimately, about the more basic question of what *naturalness* of properties is – particularly, whether one set of properties can fulfill all of the roles that natural properties have been called upon to play. It is at bottom a debate about what makes a property natural and whether special science properties are natural in any sense.

So what does it mean to say that a property (event, state, or process) is natural? Properties in one sense are easy to come by. Assuming an ontology of merely possible objects, any set of possible objects, or any function from possible worlds to extensions, will correspond to a property, according to this theory of *abundant* properties. For example, the set containing the golden mountain, Bob Dylan, my refrigerator, and the Atlantic Ocean is such a property; *grue*[60] and *bleen* are properties in this sense, as are the properties of *being an incar* (Hirsch 1980) and *being a klable* (Shoemaker 1979).

Clearly, abundant properties will not be appropriate for many of the roles for which properties are called upon to serve, for example, accounting for similarity

[60] Where something is *grue* if and only if it is green and first observed before, say, January 1, 3006, or it is blue and not first observed before that time.

between objects and contributing causal powers. To arrive at something to serve these roles, one might propose singling out a special subset of the abundant properties, the *natural* ones. Many claims have been made regarding what distinguishes natural properties from unnatural ones.[61] For instance, Lewis writes the following about natural properties: "Sharing of them makes for qualitative similarity, they carve at the joints, they are intrinsic, they are highly specific,[62] the sets of their instances are *ipso facto* not entirely miscellaneous, there are only just enough of them to characterise things completely and without redundancy. […] What physics has undertaken, whether or not ours is a world where the undertaking will succeed, is an inventory of the [natural] properties of this-worldly things." (1986, 60). "Natural properties would be the ones whose sharing makes for resemblance, and the ones relevant to causal powers" (1983, 13).[63]

These passages suggest several criteria that a property must meet in order for it to be natural:

> *Resemblance*: only natural properties underwrite similarity between objects.
>
> *Causal powers*: only natural properties contribute causal powers.

---

[61] Given the particularly unnatural, gruesome status of the non-natural properties, many philosophers have thought that they are not even properties at all. That is, they argue for a sparse conception of properties, as opposed to the abundant conception sketched above. If one adopts a sparse conception of properties, according to which all properties are natural, then the distinction can be captured by saying that the reductive physicalists claim that there are no unreduced mental properties (that are not also fundamental/restricted-sense physical properties). Note that others (David Lewis, Jonathan Schaffer) often uses "sparse" and "natural" interchangeably to refer to the privileged minority of abundant properties. I depart from this usage and use "sparse" and "abundant" to refer to two competing theories of properties. According to a sparse theory of properties, all properties are natural; there are no unnatural properties. According the abundant theory, the terms "natural" and "unnatural" distinguish between different classes of the properties that exist.

[62] This, along with the following discussion, suggests that Lewis has in mind particular properties (or perhaps magnitudes) that are only possessed by fundamental physical entities (e.g. very small masses, the quark "colors").

[63] I will set aside the complication that naturalness comes in degrees, according to Lewis, as well as the problem of how to define degrees of naturalness. With this added complication, the distinction between non-reductive and reductive physicalism amounts to a dispute about whether or not non-restricted-sense physical properties are ever *sufficiently* natural.

> *Completeness*: the set of natural properties provides the minimal basis with which to characterize the world completely and without redundancy.[64]

However, as I argued in Chapter 2, there are two senses of completeness – two things that can be meant by "characterizing things completely and without redundancy." Consequently, I claim there are potentially two sets of natural properties: natural_{FF} properties, the set of which *fundamental force completeness* is minimally true and, natural_{CS} properties, the set of which *causal sufficiency completeness* is minimally true.

Reductive physicalists claim that there is a single set of natural properties. They hold that the restricted-sense physical properties are both natural_{FF} and natural_{CS} and that mental and other broadly physical properties are neither natural_{FF} nor natural_{CS} and are thus completely redundant. Indeed, Lewis sometimes uses "fundamental" and "perfectly natural" interchangeably (e.g., 1994, 291; 1986, 60).[65] According to non-reductive physicalists, inclusive-sense physical properties, including

---

[64] Jonathan Schaffer (2004) identifies three similar "qualifications for the office of [natural] property": similarity, causality, and minimality, the last of which is the claim that "[natural] properties serve as a minimal ontological base." Schaffer argues that the minimality constraint should be abandoned and replaced with a *primacy* role; where the primary/derivative contrast corresponds to the difference between the ontological structure of reality (what is primarily real) and the linguistic truths (what is derivative or projected). If this is done, Schaffer argues that properties drawn from "all levels of nature" are primarily real from the start, and thus that the scientific properties are the only natural ones because they best fill *all* of the requisite roles. I think that this move is analogous to defining physical properties as those that are needed in an ideal theory of the world, i.e. one that runs afoul of the vacuity/unacceptable extension horn of Hempel's dilemma. It threatens to make the distinction between natural and unnatural properties vacuous. For example, does this view hold that *being grue* is natural? Or that *being an incar* or *being a klable* are? If we ignore considerations of redundancy as Schaffer (2004, 100) claims we should, aren't they on the ontological side of things, not the linguistic side, just as much as neurons and beliefs are? It seems that they are and that Schaffer's emendation lets too many properties count as natural. As discussed in the text, I do not think that the minimality/completeness role should be replaced; rather, it needs to be disambiguated.

[65] Most reductive physicalists want to count properties that are sufficiently "close" to the fundamental ones as natural. For instance, in debates about the mind-body problem it is often assumed that chemical and physiological properties are natural physical properties, even though they neither appear in basic physical theory nor are definable in basic physical terms. Whether they can consistently do this is an important question and is related to question of whether the exclusion argument generalizes, discussed in Chapter 2.

mental properties, are natural$_{CS}$ since *causal sufficiency completeness* is minimally true of the set of inclusive-sense physical properties. However, they agree with reductive physicalists that only restricted-sense physical properties are natural$_{FF}$. Again, I think that this latter claim is needed in order to ensure that their theory is a substantive version of physicalism and is not indistinguishable from (or a terminological variant of) emergentism or dualism. That is, we need the fact that restricted-sense physical properties provide a complete inventory of all fundamental forces in order to guarantee that the powers contributed uniquely by broadly physical properties do not involve any novel fundamental forces.[66]

Further, both versions of physicalism agree that the set of inclusive-sense physical properties supervenes on the set of restricted-sense physical properties. If the distribution of restricted-sense physical properties is fixed, then this fixes the distribution of the inclusive-sense properties. This of course implies that distribution of causal powers is also fixed by the distribution of restricted-sense physical properties. However, it does not imply that these restricted-sense physical property instantiations individually contribute each and every causal power. That is, it does not imply that the broadly physical properties are not needed for a complete inventory of the causal powers. If non-reductive physicalists are correct, their inventory of the world contains no redundancies with respect to fundamental forces or causal powers. Put another way, non-reductive physicalists can endorse the claim that fundamental physics is in the business of compiling a list of "the ultimate and irreducible properties of things" (Fodor 1987, 97; see also Lewis 1986, 60 quoted above and Lewis 1983, 27). Broadly physical, including mental, properties are derivative in that they are determined by (realized by) fundamental, restricted-sense physical properties.

[66] Proposals like Schaffer's (see note 64) that opt to abandon the completeness or minimality constraint are consequently not clearly physicalist theories. See also Burge (1993) and Baker (1993).

However, broadly physical properties are still irreducible in that they (a) contribute causal powers not contributed by restricted-sense physical properties *and* (b) carve up the space of causal powers in ways that restricted-sense physical properties do not.

I in effect discuss (b) further in Chapters 5 and 6. To anticipate, leaving mental and some other broadly physical properties out of an account of the world results in missed causal mechanisms, an incomplete catalog of the world's causal structure. Note that some broadly physical properties, like *being a household blender* meet (a) but not (b), if we suppose, as is plausible, that there are no scientific generalizations about household blenders as such. I think that non-reductive physicalism would not be vindicated if mental properties turned out to be "natural" only in this very weak sense.

The view that mental properties are not natural is common to any version of reductive physicalism, whether it is formulated in terms of type-identity, supervenience, or realization, and whether it holds that mental properties and events are simply ineligible to be causes or are causally heterogeneous disjunctions of natural properties that consequently do not carve out the natural, causal joints of the world.[67] I close this chapter with a discussion of causal processes or mechanisms which will be important in Chapters 5 and 6.

3.6. Causal Processes, Mechanisms and Causal Models

Mechanisms appear in many discussions of causation, explanation, and physicalism without receiving much explicit attention or clarification. To cite just a few examples: J.L. Mackie mentions causal mechanisms in his work on causation, defining one as "some continuous process connecting the antecedent in an observed ...

[67] I discuss some of these reductive versions of physicalism in Chapter 6 in the context of the multiple realization argument.

regularity with the consequent" (1974, 82). Wesley Salmon has defended a "causal-mechanical" account of explanation. According to Salmon, mechanisms are composed of causal processes and interactions, which Salmon in turn characterizes counterfactually (1984) or, more recently, in terms of conserved quantities (1994). Finally, mechanisms figure in Peter Smith's "modest" physicalist principle: "a principle P, to the effect that the behaviour of wholes is in general causally produced by the behaviour of the parts so that our explanatory stories about wholes must be consonant with our stories about the causal mechanisms constituted by their parts" (1992, 25). But Smith says nothing more about what mechanisms are or how they are constituted by parts.

Recently, mechanisms have been the subject of much philosophical scrutiny and discussion, and several competing analyses have appeared in the literature (e.g., Machamer, Darden and Craver (2000), Glennan (2002), Tabery (2004) which attempts to synthesize the first two accounts, and Woodward (2002)). The rough idea that is common to all of these views is that a mechanism is a system of entities interacting in a regular way so that changes in some entities cause changes in others, which results in a certain end condition, the exercise of a certain function, or the achievement of a goal. To describe a mechanism is to explain a phenomenon by providing details about how a certain causal process works (see Machamer, Darden and Craver 2000, 2-3). Mechanisms are especially prominent in the biological sciences; examples of biological mechanisms include cellular respiration, photosynthesis, chemical neurotransmission, and DNA replication.

Machamer et. al. (2000) give a "dualistic" analysis of mechanisms in terms of entities and activities. It is not clear to me exactly what activities are – for instance, how they differ from objects propertied and related to one another in ways that change over time. They write that "activities are the producers of change" (ibid., 3) but this

seems wrong. I think it is more plausible to say that activities simply *are* certain changes. In fact, Machamer et. al. seem to endorse this claim when they write that entities "engage" in activities (ibid., 3), since it does not seem right to say that entities engage in the producers of change. Rather, they engage in the changes themselves. Further, I have doubts about how Machamer, Darden and Craver's account of mechanisms could figure in a general account of causation or causal explanation. Machamer, Darden and Craver claim that explanation proceeds by portraying mechanisms "in terms of a field's bottom out entities and activities" (ibid., 21). Since the "bottom out" entities and activities are the "components that are accepted as relatively fundamental or taken to be unproblematic for the purposes of a given scientist, research group, or field" (ibid., 13), they thus adopt a discipline-specific view of causal, mechanistic explanation. They adopt something like Nancy Cartwright's view according to which there is no univocal, general account of causation or mechanisms, but only a variety of "thick" causal concepts that are applicable only in certain specific, localized systems and situations.[68] I think that mechanisms will likely cut across disciplinary boundaries (largely because of a phenomenon I call multiple determinativity, discussed in the Chapter 5). Consequently, I suspect that any discipline- or field-specific view of mechanisms will be inadequate.

Perhaps this can be partially remedied by noting that mechanisms can be specified at different levels of detail and will often have a hierarchical structure (Craver 2001). In addition to specifying the components of a mechanism at a single level of description, one can also describe how these component phenomena are themselves implemented by mechanisms at a lower compositional level. A fully

[68] Cartwright herself sees her project as sharing "a lot in common" with Machamer, Darden and Craver's work. See Cartwright (2004, 805).

specified, accurate mechanism at one level can be decomposed into mechanisms at a lower compositional level.

Several influential works in the causal modeling (Bayes' net/directed acyclic graph) literature also invoke mechanisms. For example, in Judea Pearl's work, mechanisms are taken to be undefined primitives (e.g., Pearl 2000). In Woodward (2003) they are characterized in terms of causal interventions, which are in turn analyzed using certain kinds of counterfactuals. According to Woodward, every structural equation in a system "represents the operation of a distinct causal mechanism" (2003, 48); mechanisms are represented by the different individual paths in a causal graph that are directed into a given variable.

Despite my misgivings about some aspects of Machamer, Darden and Craver's view of mechanisms, I wish to remain fairly neutral regarding the correct account of mechanisms. I follow Woodward to the extent that I think that most mechanisms can at least be *represented* by causal models,[69] but I incorporate the idea that mechanisms at different levels of detail can be used to explain the same phenomenon. The phenomenon to be explained by a mechanism can be represented by a single, unanalyzed link in a causal graph with a variable (or property) representing the initial conditions as cause and one representing the final condition as effect. Providing a mechanism that is responsible for this phenomenon amounts to filling in the intervening variables (properties) that lead to the production of the effect specified by the phenomenon. Once an accurate mechanism is specified at a given level of compositional detail, one can then investigate how the phenomena in this mechanism are implemented by mechanisms at lower compositional levels.[70]

---

[69] However, I do not assume that mechanisms are "modular" – roughly, that one can interfere with any given mechanism without interfering with any other (see, e.g. Woodward 2003 and Cartwright 2004).

[70] Of course, this process need not, and generally will not, proceed uniformly. One might have only a rough sketch of a complete cognitive mechanism but know in great detail how certain components of that mechanism are neurally implemented. As I discuss in the next chapter, I think that important

3.7. Concluding Remark

As mentioned above, there is one way in which this two-part argument for physicalism is programmatic. This argument assumes that the determination relation between mental and physical properties preserves inclusive-sense physicality. The argument for causal monism only establishes that the mental and physical are not independent of one another. It does not establish that the mental is dependent on the restricted-sense physical, and not vice versa. The arguments involving *fundamental force completeness* and *aggregational completeness* tell us the restricted-sense physical is fundamental, that broad-sense physical entities contribute no fundamental vital or mental forces and that, at some point in the decomposition of living and sentient entities, the parts which decomposition has yielded possess no vital or mental properties. But these arguments tell us no more than this; they leave us without a positive account of the specific way in which broad-sense physical entities are determined by restricted-sense physical ones. They assume that the determination (aggregation) relations between the physical and the mental are incompatible with strong emergentism. Providing such an account of the relation between broadly physical (including mental) entities and restricted-sense physical entities is an essential part of the non-reductive physicalist project. This is the main topic of Chapters 4 and 5.

---

debates about reduction can be construed as debates about how this process will proceed – especially whether components of a mechanism at one level of detail will line up with natural components at another level.

CHAPTER 4

VARIETIES OF REALIZATION

In Chapter 3, I claimed that one needs a positive account of the specific way in which broad-sense physical properties (including mental properties) are determined by restricted-sense physical properties. First, this account is needed to spell out how a broad-sense physical property and a restricted-sense physical property that determines it can be part of the same token causal process. Second, I claimed that an account of this determination relation should be able to answer the reductive worries that lie behind the explanatory exclusion problem outlined near the end of Chapter 3. Third, this account must be incompatible with strong emergentism regarding mental properties; it must establish that mental properties are (inclusive-sense) physical properties.

For over forty years, many non-reductive physicalists have thought that *realization* is the determination relation between restricted-sense physical properties and mental (and other broadly physical) properties. Despite this fairly longstanding position in analytic philosophy of mind, and the considerable attention that has been paid to realization in the past decade or so, there is still no consensus on the nature of realization – what it means to say that one property is realized by another. I suggest that the widespread disagreement about realization is partially explained by the fact that there are a variety of relations that have been called "realization," not all of which will meet the demands outlined above.

I explain why supervenience cannot be used as the relation upon which to build non-reductive physicalism in Section 4.1. First, as others have alleged, supervenience fails to locate, or adequately ground, mental properties in a purely physical world.

They do not meet what I will call the location constraint. Second, there need not be a token causal process in which both supervenient properties and their *distinct* base properties are involved. That is, they do not meet what I will call the causal process constraint. Indeed, supervenient properties need not be involved in any causal process. Supervenience thus cannot provide an adequate solution to the explanatory and causal exclusion problems.

Then, in Section 4.2, I undertake a survey of the seminal literature on the realization relation, focusing on how it fits into functionalist theories of mind from the 1960s, '70s and '80s. I outline the systematic ambiguity in Hilary Putnam's use of 'realization' in his papers on functionalism from the 1960s. Once-standard, functional role accounts of realization, hold that a realized property is second-order – the property of having some property or other that plays a certain role. These accounts combine a functionalist account of mental properties with a purely functionalist account of realization: the same functional role defines both the realized property and the realizer. This results in a realization relation that satisfies the location constraint at the expense of the causal process constraint, for the realizer and realized properties are no longer distinct.

Recent accounts of realization no longer rely on functional roles to do double duty in both defining realized properties and specifying how they are realized. Instead, realization is spelled out in terms of a relation obtaining between the causal powers contributed by mental properties and the causal powers contributed by their distinct realizers. Two relations have been proposed: (i) that the set of causal powers contributed by a realized property is a *subset* of the set contributed by its realizer, where the realizer and realized properties both belong to the same object;[71] (ii) that the

[71] The realizer property must also not be a conjunctive property with the realized property as one of its conjuncts.

causal profile of a realized property is *isomorphic* to the causal profile of its realizer, where the realizer is a microphysical state of affairs. I argue that the subset account can satisfy the causal process constraint, but it is not clear if it fully meets the location constraint. By contrast, the *isomorphism* account easily handles the location constraint, but arguably has trouble with the causal process constraint. Thus, these two accounts appear to do best when working in tandem; together they come closer than any other proposal in the literature to providing an adequate account of realization. However, they seem to face a dilemma – to be caught again between the inadequate physical grounding of mental properties and the reduction of those properties. In Chapters 5 and 6, I elaborate on some implicit assumptions made by these accounts to show how to avoid that dilemma.

4.1. Why Supervenience Fails: The Location and Causal Process Constraints

Robert Stalnaker (1996) writes of an elusive distinction between two intuitive interpretations of supervenience. On the one hand, some supervenience theorists (like Lewis and, in some moods, Kim) are motivated by the idea that supervenience is a liberal version of reductionism. If the properties in some set, A, are supervenient on the properties in set B, then according to the reductionist interpretation, the A-properties are completely redundant; they are "nothing over and above" the B-properties. As Stalnaker puts it, "To state an A-fact, or ascribe an A-property, is to describe the same reality [the reality that is fully determined by the B-properties] in a different way, at a different level of abstraction, by carving the same world at different joints" (1996, 222).[72] Supervenience is more liberal than traditional, logical empiricist

[72] I do not think that this adequately captures the difference between reductive and non-reductive interpretations of supervenience. For as we shall see the non-reductionist might accept one interpretation of this claim. It depends on what is meant by "determination" and "carving the same world at different joints" – whether the joints that are carved out by supervenient properties are just as natural as those carved out by the base properties (or whether they are somehow less natural, projected by us, merely pragmatic etc.).

statements of reductionism, since it is designed, as Stalnaker puts it "to isolate the metaphysical part of a reductionist claim—to separate it from claims about the conceptual resources and explicit expressive power of theories we use to describe the world" (1996, 224).

On the other hand, some supervenience theses, such as G. E. Moore's (1903) commitment to non-natural moral facts supervening on natural facts and the British emergentists views (see Horgan (1993) and McLaughlin (1992) for a discussion), are not felicitously thought of as even a liberal reductionism. Giving a precise characterization of this interpretation of supervenience is also difficult. Stalnaker describes it as one according to which the supervenience relation is "synthetic and substantive," "a sui generis metaphysical relation between distinct families of properties" (1996, 225), or "some kind of substantive metaphysical dependence" (ibid., 230).[73]

While these descriptions seem to me to be on the right track, I think that invoking the distinction between natural and unnatural properties will help to make the distinction more precise. The reductive interpretation of supervenience claims that supervenient "properties" are unnatural. According to this interpretation, supervenient properties are more abstract, more disjunctive and more unruly, less concrete and less local than the base properties and so are less suitable as causes and are less eligible for use in causal explanations. They are merely a pattern overlaid on a causal medium. The non-reductive interpretation denies this general claim. According to the non-reductive view, many supervenient properties (importantly, those that studied by the

[73] Following Yablo, one might think that there are two different versions of the non-reductive interpretation of supervenience: (1) the *emergence* interpretation, according to which the base properties are metaphysically prior to the supervenient properties and bring them into being by a kind of "supercausation", and (2) the *immanent* interpretation, according to which the base properties are inherently conditions for the supervenient properties (Yablo 1992, 256-7, n. 29). While I'm not sure that I have a firm grip on the distinction, I think that only the immanent interpretation can figure in a physicalist theory, as will become clear below.

special sciences) are natural in that they are needed in a minimal set of properties that satisfies *causal sufficiency completeness*.

With this distinction in hand, consider the arguments of a growing number of philosophers who have claimed that supervenience-based formulations of physicalism are inadequate. These critiques point out that the claim that a given set of properties, A, supervenes on another set of properties, B, merely gives us necessary covariation between the members of these sets, which is alleged to be compatible with certain forms of dualism, epiphenomenalism, and any view according to which the A-properties are non-physical but are in pre-established harmony with the B-properties.[74] The general thrust of these arguments is that supervenience is not strong enough to guarantee that the supervenient properties are "physicalistically acceptable," (that is, "broadly physical" in my terms) even if *metaphysical*[75] necessity appears in the definition of strong supervenience. Any supervenience-based account of physicalism is compatible with views that are competitors to physicalism (or immediately turn the dispute between rival theories into an undecidable clash of intuitions) (see Wilson 2005)), e.g., occasionalist forms of dualism where the metaphysically necessary supervenience of the mental on the physical is established by God's decree. Even assuming that the claim of mental-physical supervenience is sufficient for the dependence of the mental on the physical (and is not merely a covariance claim), the thought is that this supervenience claim leaves out what *grounds* or *accounts* for this dependence (see, e.g., Kim 1998, 13). Any number of physically unacceptable

[74] See, e.g., Horgan (1993), Kim (1998, ch.1), Hawthorne (2002), Melnyk (2003, ch. 2), and Jessica Wilson (1999, 2005) for further discussion of the weaknesses of using supervenience to formulate physicalist theses. Bailey (1999) and Noordhof (2003) defend supervenience-based formulations of physicalism.

[75] What about using logical supervenience in a formulation of physicalism (a la Chalmers 1996)? This is plausibly not vulnerable to the counterexamples that have been put against metaphysical supervenience. However, this is arguably because logical supervenience is associated with claims of *a priori* entailment. Thus, it is not supervenience that is doing the work in the account of physicalism, but rather conceptual links between the supervenient and base properties.

grounds is compatible with the supervenience claim, e.g., a god's decree or an emergent fundamental force that comes into existence only in sentient creatures.[76]

Many of the critics would go on to say that supervenience-based formulations fail because they are not *explanatory*. For instance, they fail to give an account of how mental properties are determined by the properties of the brain (and perhaps the environment). To posit a *sui generis* relation between mental and physical properties and then claim that this secures the place for the mental in the physical world is "obscurantist" (as Stephen Schiffer (1987), one of the earliest such critics, claims).

I think it is clear that these criticisms are non-starters if we adopt the reductive reading of supervenience; if the criticisms are sound, they hold *only* for the non-reductive interpretation. According to the reductive interpretation, ascribing supervenient "properties" is *merely* a different way of talking about a reality that is determined by the set of fundamental, natural properties. The supervenient "properties," according to the reductionist, are *completely* redundant—unneeded in an account of what there is—because they are unnatural. If one adopts a sparse theory of properties, then according to the reductionist, there are no supervenient mental properties. There are only mental concepts or property designators. But if this is so, then there does not seem to be any obscure supervenience relation, between distinct sets of natural properties, to clear up.[77] In other words, one cannot set up the kind of counterexample where the metaphysical dependence of the supervenient on the base properties is established in some external or deviant way (God's decree, an emergent fundamental force). For, in such cases, the supervenient domain is *not* just a way of

[76] This criticism exploits the following formal fact: if a set of properties C (e.g. mental properties) supervenes on the set (A ∪ B) (e.g. physical properties and properties of God) and there is a perfect correlation between instantiations of properties in A and B, then C also supervenes on A.

[77] Cf. Hare's (1952) view that supervenience is a conceptual constraint on moral discourse. For Hare, the claim that the moral supervenes on the physical/natural does mean that a set of moral properties supervenes on a distinct set of natural/physical properties. Hare, of course, denied the existence of moral properties.

"carving the same world at different (less natural) joints." In addition to the reality determined by the base properties, there is also a "metaphysical dangler"—a condition that is not itself a base property but that is required for the appropriate patterns of instantiation of the supervenient properties to obtain.

However, if one adopts the non-reductive interpretation of supervenience and wants to remain a physicalist, then one needs to ground or account for this apparently brute metaphysical supervenience relation. That is, one has to ensure that the dependence (determination) results from the nature of the base properties alone and does not arise from some external or "deviant" source. Realization has been proposed as a promising candidate for performing this function, that is, as a relation that is stronger than (i.e. entails but is not entailed by) supervenience and provides for an explanation of how the supervenient properties depend on (are determined by) the base properties in a physicalistically acceptable way. That is, realization was offered as a replacement for supervenience – as a relation such that if anything bears it to (restricted-sense) physical properties, then that entity is *thereby* (inclusive-sense) physical.

In terms of offering a solution to exclusion problem, supervenience fails because it cannot be used to explain how there is some causal process in which both supervenient properties and their base properties play a part. (Indeed, supervenient properties need not play a role in any causal process.) That is, supervenience fails to satisfy the following closely related constraints:

> *Location constraint*: the relation between mental (and other broadly physical) properties and restricted-sense physical properties must provide for an account of why mental properties are inclusive-sense physical, that is, it must provide for a physicalistically acceptable account of how mental (and other broadly physical) properties are grounded in restricted-sense physical properties.

> *Causal process constraint*: the relation between mental (and other broadly physical) properties and restricted-sense physical properties must provide for an account of (a) how the mental properties can be part of the same token causal process as restricted-sense physical properties but (b) remain distinct.

The location constraint amounts to locating the mind in the physical world – to spelling out how mental properties are determined by the properties of restricted-sense physical entities. The causal process constraint of course arises from the need for a solution to the exclusion problem (regarding causal sufficiency) and the subsequent exclusion-based worries about causal explanation. In the next section, I turn to the realization relation and the question of whether it can succeed in meeting these constraints.

4.2. Realization in Putnam's Work: Functional Role and Natural Property Realization

Hilary Putnam introduced the notion of realization in a series of six papers written from 1960 to 1973. These discussions of realization used analogies and ideas from the relatively new discipline of computer science. Mental states were thought to be importantly similar to the computational states of machines. Just as the same program or software could be implemented or realized in systems composed of different physical objects (vacuum tubes, transistors, etc.), so, the idea was, a given mental state could be realized in physical states or substances that had nothing important in common, aside from their all being integrated into a system so as to realize the given mental state. As Putnam so colorfully put it: "We could be made of Swiss cheese and it wouldn't matter" (1975b, 291).[78]

Putnam relied on the analogy to Turing machines to provide content to the claim that mental states are realized by neurological states. However, this analogy

[78] I think that it wouldn't matter only assuming (probably contrary to fact) that Swiss cheese can enter into all of the causal (and not merely computational) processes that neurons do.

was not rich or specific enough to yield a fully worked out theory of realization. In Putnam's papers, several distinct ideas are combined in his intuitive use of realization; the notion of realization in his work is multiply ambiguous.

This ambiguity is reflected in the many different kinds of entity that are said to bear the realization relation to each other. Among the entities which Putnam claims are realized are: Turing (or abstract) machines (1960, 371), logical and mental states (1960, 372-3), functional organizations (by which he seems to mean, an inductive logic or a rational preference function) (1963, 327), psychological predicates (1964, 390), agents (1967a, 414), machine tables (ibid.), sense organs (1967b, 434), sensory inputs (ibid.), and abstract structures (i.e. functional/psychological theories) (1975b, 294). Among the entities said to be "realizations" (i.e. what I call "realizers") are: structural, physical, or internal states (1960, 372-3; 1964, 39), physical attributes/predicates (1964, 392), automata built out of various materials (1967a, 414), empirically given systems (1967b, 434), and "realizations" or "ways" which were not further specified (1960, 371; 1967a, 416; 1967b, 434, 438).

I think that two main ideas can be filtered out of this confusing welter, which are illustrated by the following passages:

> … the 'logical description' of a Turing machine does not include any specification of the physical nature of these 'states' – or indeed, of the physical nature of the whole machine. (Does it consist of electronic relays, of cardboard, of human clerks sitting at desks, or what?) In other words, a given 'Turing machine' is an abstract machine which may be physically realized in an almost infinite number of different ways.
>
> As soon as a Turing machine is physically realized, however, something interesting happens. Although the machine has from the logician's point of view only the states A, B, C, etc., it has from the engineer's point of view an almost infinite number of additional 'states' (though not in the same sense of 'state' – we shall call these structural states). For instance, if the machine consists of vacuum tubes, one of the things that may happen is that one of its vacuum tubes may fail – this puts the machine in what is from the physicist's if not the logician's point of view a different 'state'. Again, if the machine is a

> manually operated one built of cardboard, one of its possible 'non-logical' or 'structural' states is obviously that its cardboard may buckle. And so on. (Putnam 1960, 371)
>
> Psychological attributes, whether in human language or in robot language, are simply not the same as physical attributes. To say that a robot is angry (or 'angry') is quite a different predication from the predication 'such and such a fluid has reached a high concentration', even if the latter predicate 'physically realizes' the former. Psychological theories say that an organism has certain states which are not specified in 'physical' terms, but which are taken as primitive. … Thus, as Jerry Fodor has remarked … it is part of the 'logic' of psychological theories that (physically) different structures may obey (or be 'models' of) the same psychological theory. … the pattern of correct usage, in the case of an ordinary-language psychological term, no more presuppose or imply that there is an independently specifiable state which 'realizes' the predicate, or if there is one, that is a physical state in the narrow sense (definable in terms of the vocabulary of present-day physics), or, if there is one, that is the same for all members of the speech community, than the postulates of psychological theory do. (Putnam 1964, 392)
>
> The Turing machines I want to consider will differ from the abstract Turing Machines considered in logical theory in that we will consider them to be equipped with sense organs by means of which they can scan their environment, and with suitable motor organs which they are capable of controlling. (Putnam 1967a, 409)
>
> Note that a Turing Machines need not even be a machine. A Turing Machine might very well be a biological organism. … Strictly speaking, a Turing Machine need not even be a physical system; anything capable of going through a succession of states in time can be a Turing Machine. (ibid., 412)
>
> There is, however, a sense in which we may say of these agents [i.e. Turing machines], regardless of their physical realization, that they are conscious of certain things and not conscious of others. (ibid., 414)

In the first passage, the focus is on the abstract, *logical* states of Turing machines; realization captures the idea that the logician can rightly ignore some of physical details of the machine that implements the algorithm represented by the machine table. These physical details are "solely the concern of engineers," as

Putnam puts it.[79] The second passage ties realization to *psychological theories* and the pattern of correct usage of ordinary-language psychological terms. According to Putnam, neither psychology nor common sense presupposes anything about the realization, physical or otherwise, of mental predicates. In the final passage, the emphasis changes. Instead of focusing on the realization of logical or functional structures (represented by psychological theories), the emphasis is on the realization of empirical entities – biological or mechanical (or perhaps even non-physical) systems with sense organs.

As these passages demonstrate, Putnam uses the term 'realization' to stand for (at least) two different sorts of relation:

(1) *functional role realization*, where the realized entities are "second-order properties" – properties that are defined by conditions or roles that are satisfied by other (first-order) properties. This relation falls into subtypes which are distinguished by the notion of function that is invoked:

> (1a) *computational role realization*: a relation between a formal or mathematical structure or property (e.g. what a Turing machine table specifies) and a physical system or property that implements the former.
>
> (1b) *theoretical role realization*: a relation that holds between a second-order property or predicate (property designator) and a physical property if and only if the physical property is part of the domain of a model for a scientific or folk theory that is concerned with the second-order property (contains the second-order predicate). The theoretical role associated with the second-order property is usually claimed to be delivered by conceptual analysis.

[79] Alan Turing made a similar point in a BBC radio program broadcast on 10 January 1952: "The important thing is to try and draw a line between the properties of a brain, or of a man, that we want to discuss, and those that we don't. To take an extreme case, we are not interested in the fact that the brain has the consistency of cold porridge. We don't want to say: 'This machine's quite hard, so it isn't a brain, and so it can't think'" (Turing 2004, 494-5).

(1c) *causal role realization*: a relation between a second-order property and a physical property, where the physical property plays the causal role that is definitive of the second-order property.[80]

(2) *natural property realization*: a relation between two natural states or properties, usually characterized in terms of the causal powers contributed by each.

The ambiguities in Putnam's understanding of realization roughly parallel ambiguities in his use of the concept of a Turing machine. Early on, he thought of Turing machines as abstract mathematical entities, as had become customary in computability theory. But in the third passage above, from the "The Mental Life of Some Machines" (1967a), he explicitly introduces a new notion of Turing machine. In contrast to the abstract Turing machines studied in logical theory, these Turing machines have sense and motor organs with which they can interact with their environment. In worlds in which physicalism is true, such machines will be complicated physical devices or systems.[81] Instead of only formal or mathematical entities being realized, Putnam now claims that agents and robots (understood as the new Turing machines), their sense and motor organs, and sensory inputs are *themselves* realized in different ways. The important thing to note here is that the realized entities (as well as the realizers, of course) include things that are empirical and spatio-temporal, and which are plausibly causally efficacious. They will also both be physical (in the inclusive sense), if physicalism is true in the actual world.

However, a few pages after introducing this notion, Putnam claims that Turing machines need not even be physical and that "anything capable of going through a succession of states in time can be a Turing Machine" (1967a, 416). This introduces another notion of Turing machine. According to this interpretation, something is a

[80] This leaves out biological or teleological notions of function which some have argued are crucial for rebutting objections to functionalism about the mind (see Lycan (1987) and Sober (1985)). An adequate discussion of teleology is beyond the scope of this dissertation.
[81] Newell and Simon's (1976) notion of a "physical symbol system" is similar to this notion of a Turing machine and fits into their view of computer science as *empirical* inquiry.

Turing machine just in case (and because) it can be described at a certain level of abstraction. It is natural to think that on this conception, 'Turing machine' is primarily a descriptive notion. Turing machines are not abstract mathematical entities, nor are they (inclusive-sense) physical entities that are realized by other physical entities. Rather, all that exists, on this view, are restricted-sense physical entities, configurations of which can be described at a variety of levels of description or abstraction. When an aggregate can be described at the appropriate level of description (computational, functional, semantic, or whatever), then it (that aggregate) is thereby a Turing machine. There is obviously a natural affinity between the versions of realization introduced above and these interpretations of Turing machines.

4.3. Functional Role Realization

As noted, Putnam offers no explicit account of realization in any of his papers; it is tacitly defined in terms of notions like "sameness of functional organization" and "functional isomorphism." But, of course, the notion of a function is itself multiply ambiguous. Three core notions of function correspond to the different strands in Putnam's discussion: *computational function*, *theoretical function*, and *causal function*. Throughout the 1970s, '80s, and '90s, philosophers developed these strands into three distinct but often entwined forms of functionalism: computational functionalism; central-state identity theory or theoretical role functionalism, and causal role functionalism.[82] According to computational functionalism, realized entities are formal or mathematical – realized processes are purely mathematical and "mirror" (in a sense to be specified) the realizing physical processes (for further discussion of various forms of computational functionalism see (Piccinini 2004)).

[82] Note that this three-fold distinction cuts across the distinction between analytic functionalism and psycho-functionalism. The latter is a semantic contrast, a debate about how the meaning of functional terms is specified and what our mental concepts and terms refer to. My distinction is drawn with respect to what kind of function is involved, about the nature of the functional entities that are realized.

This relation is often discussed in the cognitive and computer sciences, where it is called "implementation." Computational versions of functionalism see the mind as literally software that the hardware of the brain circuitry implements. Theoretical role functionalism takes realization to be a semantic relation – realization is the satisfaction relation. If an entity satisfies (a formula of) a scientific or "folk" theory, i.e. if is the occupant of a theoretical functional role specified by the formula, then that entity realizes (that formula of) the theory. According to causal role functionalism, realizers play the causal role that is definitive of the realized property.

The common core of these accounts of realization is that a realized property is by definition a second-order property – a property that is possessed by an object if and only if that object possesses some other (first-order) property that satisfies a certain condition or role.[83] Such properties are second-order in that they are "generated by quantification over the base [first-order] properties" (Kim 1998, 20).

4.3.1. Computational Role Realization

As mentioned above, computational role realization has received the most thorough discussion in computer and cognitive sciences (and philosophical discussions that draw on these fields) – where it is discussed under the name 'implementation' (see, e.g., Chalmers (1994), Scheutz (1999, 2001). Computational role realization is a relation between an "abstract computation" (the realized entity) and a "physical system" in which it is realized.[84] "The causal structure of the physical system mirrors the formal structure of the computation" (Chalmers 1994, 392). This version of

[83] This should not be confused with another notion of second-order property, a property of a property. The functional role itself is a second-order property in this sense. In the functionalist literature, both first-order and second-order properties are properties of individuals. First-order properties are either "primitive" properties or those that can be defined in terms of primitive properties plus quantification over individuals. Second-order properties are properties of individuals that are definable only in terms of individuals plus quantification over individuals *and* first-order properties. For more on the functionalist framework and second-order properties, see Loar (1991) and Hill (1991, 47).

[84] Note that the term "implementation" is sometimes used loosely. Computer scientists often claim that a mathematical function (or even a piece of software) is implemented in another piece of software (e.g. neuroimaging software may be implemented in the program MATLAB).

realization can be traced back to the early days of artificial intelligence and is captured by the claim that: "Turing opened the way for people who study the *logic* of the brain without any interest in the physical matter of the support" (Hodges 1983, 291, italics added). In its simplest form, computational role realization claims that a physical system realizes a computation if and only if a one-one correspondence obtains between the set of computational states and a set of first-order, physical states of the system. Putnam subsequently argued that this account of realization was too weak. For it can be shown that according to this account of realization any physical system will realize any computation, as long as one helps oneself to a fairly liberal criterion for what counts as a first order physical state (Putnam 1988, Searle 1992). Much of the subsequent discussion in the philosophical literature tries to avoid this result by placing more restrictions either on the relation between states (demanding more than that a one-one correspondence holds) or on what counts as a physical state. For example, Scheutz (2001) suggests that the mathematical relation of "bisimulation" should be used as a replacement for correspondence- and isomorphism-based accounts, where, very roughly, bisimulation requires only a one-one correspondence between computational and physical-causal paths, not a mapping between individual states.[85]

[85] For other discussions relevant to implementation, see Melnyk (1996) and Frances Egan's (1991, 1995) mathematical "function theoretic" interpretation of David Marr's computational level. Computational role realization is not limited to discussions drawing on computer and cognitive sciences. A similar view of realization appears in Mohan Matthen and André Ariew's (2002) work on fitness and natural selection. They appeal to realization as part of their argument that natural selection is not a cause of evolutionary change but rather is merely a "mathematical trend" or "formal phenomenon." Three aspects about their view of realization make it appropriate to classify it as mathematical: (1) realizers of more abstract selection formulae are produced by the logical operation of conjunction, (2) it is solely the "formal character" of natural selection theory that accounts for the fact that natural selection is multiple realizable, and (3) what unifies all of the realizers of a single realized property is their formal, not causal, structure (see Matthen and Ariew 2002, 74-7, 71-2, 81). Daniel Dennett's (1991) discussion of the "Game of Life" seems to suggest a similar view of realization. Here the realized entities are mathematical patterns that are formal consequences of the lower-level, physical phenomenon, e.g. stable patterns produced by various initial conditions.

4.3.2. Theoretical Role Realization

David Lewis's articulation of functionalism (which is perhaps more accurately described as a version of type-identity theory) is the most fully developed account of theoretical role realization. Here the realized entity is a theory (or an open formula formed from the conjunction of all of the sentences of a theory). The procedure for showing how the theory is realized runs roughly as follows. One takes a newly formulated theory and forms a long conjunction of all of the sentences that contain newly introduced theoretical terms. Then one uniformly replaces the theoretical terms with variables to obtain an open formula. "Any n-tuple of entities that satisfies this formula, under the fixed standard interpretations of its O-terms [original or old terms], may be said to *realize*, or to be a *realization of*, the theory T" (Lewis 1970, 430). In Lewis's toy example, three men, "Plum, Peacock and Mustard together *realize* (or are a *realization of*) the detective's theory" about who committed a certain crime (Lewis 1972, 251). That is, the three people satisfy the formula constructed as outlined above; the people are in the only elements in the domain of the intended interpretation of the theory.

If a theory is uniquely realized, theoretical role realization can be used in a straightforward argument for type identity. This argument starts with the meaning of the theoretical terms, which is determined implicitly by the Ramseyfied theory. One then discovers, through empirical investigation, that the theory so interpreted is true of certain entities. So, Lewis claims, we can conclude that the theoretical terms denote these entities (if they are the unique satisfiers of the theory). Applied to common sense psychological theories, Lewis believes we can use this procedure to show that terms for mental states denote their physical realizers. Folk psychology provides us with the meanings of mental terms "by analysis": "mental state *M* = the occupant of the *M*-role. Empirical science informs us about which (hopefully) physical entity

plays that role in the actual world: "physical state *P* = the occupant of the *M*-role, therefore *M* = *P*" (Lewis 1994, 303).

If the same mental state is realized by different states in different systems, this does not tell against the identity theory, according to Lewis. Rather, "our psychophysical identities need to be restricted: not plain *M* = *P*, but *M*-in-*K* = *P*, where *K* is a kind within which *P* occupies the *M*-role" (Lewis 1994, 305). Note that these restricted identities are metaphysically necessary. *Pain-in-humans* is necessarily identical to, say, *c-fibers firing* (if that is what realizes *being in pain* in the population of humans). It is only terms like *being in pain* that are non-rigid, according to Lewis. So, *being in pain* could have been something else, just like the winning lottery number could have been different (see Lewis 1980, 233), but *pain-in-humans* is *c-fibers firing* in every possible world.

4.3.3. Causal Role Realization

Causal role functionalism was the default view of realization in the '80s and '90s, perhaps because it explicitly appeals to the functionalist contrast between roles and their occupants in defining realization. It is commonly assumed in discussion of mental properties that the condition or role of interest is a causal one. Several of the accounts that I discuss below do not explicitly mention *causal* roles, but it seems clear that this was the notion of functional role that was assumed to be most relevant to the philosophy of mind. Some of these passages could also be interpreted as articulations of (non-causal versions of) theoretical role functionalism. In this section I am just interested in the causal interpretation.

In his recent book Mind in a Physical World, Kim sketches an account of realization that relies on the notion of a second-order property. While he never explicitly defines what it is for one property to realize another, he utilizes the idea that, in general, second-order properties defined over some set of base properties are

realized by those base properties. Some examples are: the property of *having a primary color* which is realized by *being red*, *being yellow*, and *being blue*; the property of *being jade* (that is, according to Kim, the second-order property of *being a mineral that is pale green or white in color and fit for use as gemstones or for carving*) which is realized by *being jadeite* and *being nephrite*; and *dormitivity*, which is realized by chemical properties such as *being diazepam* and *being secobarbital* (1998, 20-1). Thus, according to this conception, "[M]ental properties are specified by causal roles, that is, in terms of causal relations holding for first-order physical properties (including biological and behavioral properties). … To be in a mental state is to be in a [first-order] state with such-and-such as its typical causes and such-and-such as its typical effects" (1998, 21).

Many other authors give more or less the same construal of realization.[86] Here are two recent examples. Andrew Melnyk claims that: "Token *x* realizes token *y* iff (i) *y* is a token of some functional type, F, such that, necessarily, F is tokened iff there is a token of some or other type that meets condition, C; (ii) *x* is a token of some type that in fact meets C; and (iii) the token of F whose existence is logically guaranteed by the holding of condition (ii) is numerically identical with *y*" (2003, 21). Similarly, David Papineau gives the following account of realization: "The mental fact that person X has mental propery M is realized by the physical fact that X has physical property P if and only if M is a higher-level property – a property of instantiating some lower-level property with certain features R – and P's instantiation in X has these characteristics R" (1995, 237).

[86] See also Beckermann (1992, 18), Antony and Levine (1997, 85), Francescotti (2002, 283), among many others.

4.4. Functional Role Realization and the Location and Causal Process Constraints

Each of the versions of functional role realization can easily handle the location constraint. According to these accounts, second-order realized mental properties are inclusive-sense physical because they are merely a different way of picking out or expressing (perhaps disjunctively) restricted-sense physical properties.[87]

However, the ease with which they fulfill the location constraint suggests that they cannot satisfy the causal process constraint. Computational role realization obviously does not fare well with respect to part (a) of the causal process constraint. Formal, mathematical properties, like *being an Abelian group* or the logical relations represented by a Turing machine table, are paradigmatic examples of properties that are not causally efficacious. The *formal* structure of the realized computation is isomorphic to the *causal* structure of the physical system that is the realizer. There is no isomorphism between two causal systems. Thus, according to computational role realization, no realized entities are part of any causal process. So this account certainly cannot explain how some realized broadly physical properties are involved in the same token causal process as restricted-sense physical properties.

According to theoretical role realization, realized "properties" are merely different ways of describing restricted-sense physical reality. In fact, they are not natural, causally efficacious properties at all. So this account of realization fails to meet part (b) of the causal process constraint. To think that realized "properties" like *being hot* or realized states like *pain* are causally efficacious would in Lewis's words:

---

[87] One might have nominalist worries about the physical acceptability of mathematical properties in general. If so, then one will not think that computational role realization satisfies the location constraint. However, the physicalist theses I am concerned with are neutral on the question of whether abstract mathematical entities are compatible with the thesis that everything is physical. I am concerned with physicalism's implications for the mind-body problem and the relation between various sciences, not with the nature of abstract mathematical objects and other issues associated with the debate between Platonism and nominalism about mathematical entities. Thus, I will freely refer to mathematical entities, without assuming that they are reducible to, say, space-time points.

"multiply causes beyond belief by playing a verbal trick" (1983, 44). Such properties are, according to theoretical role realization, too unnatural, too disjunctive and extrinsic, to be causally efficacious.

Proponents of theoretical role realization, in effect adopt a reductive interpretation of supervenience. That is, they hold that supervenience is not a relation between distinct sets of natural properties. The supervenient domains that are the purview of the various special sciences and folk theories are merely another way of talking about the subvenient domain of restricted-sense physical properties (the only natural properties there are). Thus, if realization's role is to explain a metaphysical supervenience relation between distinct sets of natural properties, then there is no need for realization in a reductive account of physicalism. Of course, the reductionist may still appeal to a relation called "realization" to spell out how the basic properties in fact determine what the world is like, but, such a relation will not fulfill the requirements on non-reductive versions of physicalism.

Causal role realization also fails to meet the causal process constraint, and spelling out why will be instructive when the discussion turns to finding an account of realization that does meet this constraint. Suppose that mental properties are second-order properties defined in terms of causal roles (relations between inputs, outputs, and other mental properties). Now suppose that what it is for one property to realize another is for the first to play the causal role associated with the second. But now, if we rule out massive overdetermination or double counting of causes, we cannot say that the same causal role is played twice over, once by the mental property and once by its physical realizer.[88]

[88] Note that this is not exactly same worry as that raised by the exclusion problem. Since "the completeness of physics" has been disambiguated, it is not clear that the exclusion problem can get off the ground. The problem is rather that we have a single causal role (or profile) that is allegedly defining or characterizing two properties.

The trouble that functional role realization poses for keeping mental states distinct from their realizers, while maintaining their causally efficacy, is nicely demonstrated by the following passage:

> The 'functional causal role' of pain will be the causal role assigned to pain by a maximally good functional definition. This will be different from *its* [i.e., pain's] *total causal role*, that is, the totality of its causal features (otherwise it would not be possible, as functionalists insist it is, for pain to be physically realized in different ways – *each possible realization will correspond to a total causal role*, and the functional causal role will be what all these have in common). (Shoemaker 1981, 317, italics added)

Here the causal role functionalist is engaged in a kind of double-speak. If the "total causal role" corresponds to both the mental property and its physical realizer, and we accept a minimal causal theory of properties, we have just one property that is expressed in two different ways, by mental and physical predicates. If one wants to satisfy the causal process constraint, one should not define realization in terms of the realizers "playing the causal roles" of the properties they realize since this immediately results in there being only one property involved in the given causal process, thus violating the distinctness claim in part (b) of the causal process constraint.

The causal role account of realization fails to meet the causal process constraint because it is inextricably linked to the functionalist account of mental properties. According to the functionalist theory of mind, something has a mental property if and only if it has some other property that plays a certain causal role. But this just *is* the causal functional role account of realization. Together, the account of the nature of mental properties and the account of how they are realized result in realized mental properties being identical to their restricted-sense physical realizers

(bracketing the multiple realization of mental properties).[89] A single causal role is ascribed to the realizer and realized property, and a weak causal theory of properties that is common ground in this debate ensures that the mental "property" is simply another way of describing the restricted-sense physical property.[90]

This suggests that we can obtain an account of realization that meets the causal process constraint by separating the account of realization from the functionalist account of mental properties. However, non-causal versions of theoretical role realization fare as well as the causal role account. Both meet the location constraint but fail to meet the causal process constraint, which is unsurprising given the similarity between causal role functionalism and theoretical role functionalism. Why not attempt to modify a non-causal version of theoretical role realization so that it meets the causal process constraint?

Non-causal versions of theoretical role functionalism are "ideological" or "conceptual" in that they in effect eliminate natural functional properties and retain only functional concepts.[91] Causal role functionalism, with its accompanying account of realization, is also ideological or conceptual in this sense, since it leads to the identification of realized properties with their realizers. Realized entities are only distinct as concepts. However, according to causal role realization, whether one property is a causal role realizer of another is partly a matter of something *independent* of any functionalist theory of mind – namely, what causal relations obtain in the world. Causal role realization is not *in principle* tied to functionalist theories of the mind. However, according to theoretical role realization, whether or not one entity is

[89] Whether the multiple realizability of mental properties puts this into doubt will be discussed in Chapter 6.
[90] Cf. Shoemaker (2001, 84, 77, 75-6) on the problem raised by trying to combine two tasks: having causal roles both define mental properties and account for how mental properties are realized.
[91] For a discussion of how Kim adopts this kind of "conceptual functionalism" see David (1997). See also Bealer (1997) for a related notion of "ideological functionalism."

(non-causally) theoretical-role realized by another is entirely a semantic or theoretical matter and is thus inextricably linked to some functionalist (folk or scientific) *theory* of mind. As a semantic relation, theoretical-role realization is inherently a relation with a linguistic/theoretical entity as one of its relata. So there is no way to modify a non-causal version of theoretical role realization so that it is independent of some functionalist theory of mind. In contrast to causal role realization, there is no independent, non-theoretic component of theoretical role realization that can be used to pry it away from functionalist theories of mental properties, concepts, or predicates. Theoretical role realization is *essentially* ideological. Consequently, any modified account of a non-causal version of theoretical role realization will be unable to meet the causal process constraint.

4.5. Causal-Explanatory Accounts of Natural Property Realization

Given that one cannot invoke functionalism to both define realization and formulate a theory of mental states, philosophers began to develop accounts of realization that were independent of functionalist theories of mind.[92] Thus, much of this recent work on realization can be seen as an attempt to correct the causal functional role account of realization so that realization meets the causal process constraint. These accounts take realization to be a relation between distinct empirical, causal properties, and they explicitly claim that realized properties contribute their own causal powers, instead of merely specifying which causal powers must be contributed by (or which role is filled by) their realizers. A forerunner to these accounts is the view that realization should be defined in part by an explanatory relation that holds between the realizer and realized entities. Many of these accounts

[92] To my knowledge, Christopher Peacocke (1979, 116) was the first to propose that "the label 'realization' should be drained of any association it may have with functionalist doctrines in the philosophy of mind." Although his work is sometimes cited in the literature on the causal argument for physicalism, it has been overlooked in contemporary discussions of realization. See n. 93 below.

are motivated by the idea, discussed above, that supervenience-based accounts of physicalism fail because they do not explain how mental properties are determined by physical ones. I discuss these explanation-based accounts of realization before turning to causal power accounts of natural property realization.

4.5.1. Explanation-based Accounts

One influential explanation-based account of realization comes from Ernest LePore and Barry Loewer who write:

> Exactly what is it for one of an event's properties to realize another? The usual conception is that e's being P realizes e's being F iff e is P and e is F and there is a strong connection of some sort between P and F. We propose to understand this connection as a necessary connection which is explanatory. The existence of an explanatory connection between the two properties is stronger than the claim that P→[F] is physically necessary since not every physically necessary connection is explanatory. For e's being P to explain its being F it may be necessary for there to be a *system* of connections between realized and realizing properties or property kinds to which P and F belong. And it may require that the central laws and principles governing the realized properties be explained by the connections between basic and non-basic properties and laws governing the basic properties. For example, systematic variations in the molecular structure of a substance give rise to systematic variations in its degree of solubility and possessing of a particular molecular structure explains the rate of dissolving and so forth. This supports the claim that a substance's molecular structure realizes its solubility. (1989, 179-80, italics in original)

Here LePore and Loewer explicitly claim that an explanatory relation is needed to strengthen the nomologically necessary connection provided by supervenience. Two other authors offer explanation-based accounts of realization that are quite similar to LePore and Loewer's. First, David Pineda claims that "there must be an explanatory link between realizers and realized properties in such a way that each instantiation of the realized property can be explained in terms of the instantiation of its realizer on that occasion" (2001, 46-7). The motivation for introducing this

explanatory constraint is the same as LePore and Loewer's – to distinguish realization from mere property correlation or supervenience.

Second, in the context of a very different project (formulating a "use-theory of meaning") Paul Horwich offers what I interpret as an explanation-based account of realization (what he calls "constitution"). He claims that: "the content of the claim that the relatively superficial property, 'being *f*', *consists in* the underlying property, 'being *g*', is to assert a fundamental explanatory relation between them. It is to say not merely that all *f*s are *g*s and vice versa, but also that this biconditional explains various facts about the superficial property" (Horwich 1998, 178-9).

Finally, although Terence Horgan, like Horwich, does not claim to be giving an account of realization, he also thinks, with LePore and Loewer and Pineda, that explanation is the key to strengthening supervenience into a relation useful for formulating physicalism. Horgan claims that a materialist who is also a realist about mental states needs a relation that he dubs 'superdupervenience': "ontological supervenience that is robustly explainable in a materialistically [acceptable] way. Superdupervenience would indeed constitute a kind of ontic determination which is itself materialistically kosher, and which thereby confers materialistic respectability on higher-order properties and facts" (1993, 566). Horgan's superdupervenience plays the same role as the explanatory version of realization, namely, as a stronger relation than supervenience which is supposed to be an essential part of formulating a materialist metaphysics. Barring other reasons to dissociate superdupervenience from realization, I will view it as within the ambit of explanation-based accounts.

While I think these accounts should be grouped together, there are significant differences between them, especially concerning the nature of the explanandum and explanans. For LePore and Loewer, Pineda and Horwich, the realization relation "does the explaining." The instantiation of the realizing property explains the

instantiation of and facts about the realized property. For Horgan, on the other hand, the explanatory relation does not hold between properties (or their instantiations), rather, it is the supervenience relation itself that must be explained. Thus, instead of explaining a particular fact (or set of facts), i.e. the instantiations of supervenient properties by events, Horgan's account requires something closer to the explanation of a law. Jaegwon Kim makes roughly this point, when he claims that merely having Nagelian bridge laws (which are all that supervenience requires) would not be enough for physicalism (Kim 1998, 95-6). As he writes, "the mere claim of mind-body supervenience leaves unaddressed the question of what *grounds* or *accounts for* it—that is, the question why the supervenience relation should hold for the mental and the physical" (1998, 13).[93]

Only a few authors explicitly invoke explanation in their definitions of realization, but nearly everyone in the debate believes that realization provides the basis for certain kinds of explanation. For example, Melynk, who is a proponent of a functional role theory of realization (discussed above), claims: "LePore and Loewer offer an apparently quite different account of realization as a 'necessary connection which is explanatory'…. But, … when the relationship between realization physicalism and reductionism is explored, realization as I understand it is intimately connected with explanation, so that their account is closer to mine than it initially appears" (2003, 21 n.17). Lenny Clapp, an advocate of the subset account (discussed below), refers to LePore and Loewer's account as a "working definition" of realization according to which the instantiation of the realizer explains "in some metaphysical sense" the instantiation of the realized property (Clapp 2001, 112-3, 129). He

[93] In a footnote, Kim (1998, 123 n.21) cites Horgan as the first to note the need for a physicalistically acceptable explanation of supervenience (see Horgan (1993, 578) and the citations therein).

apparently believes that his own proposal about realization fleshes out this working definition.

I think that it is unnecessary to treat the explanation-based accounts separately. I agree with Kim's "explanatory realism" which claims that "we should look for an objective metaphysical relation between [the realizer] and [realized property], not an essentially epistemic relation like explanation; that is, we should view the explanatory relation between the two properties as being supported by a metaphysical realization relation" (1993a, 343). If this perspective is adopted, then the explanation-based accounts are not genuine alternatives to varieties of natural property realization; rather, they are more superficial or programmatic accounts natural property realization. They collapse into some causal power account of natural property realization, to which I now turn.

4.5.2. Causal Power Accounts

4.5.2.1. The Subset Account

The causal power account is foreshadowed by some functional role accounts.[94] For example, according to Beckermann: "If a system S is in a (mental) state F at time t, F can be said to be realized at t by the (physical) state G if and only if S is in G at t and G has in S all the (monadic, relational, etc.) features that are characteristic for states of kind F" (1992, 18). This of course leaves open the possibility that the realizing state possesses features which the realized state does not. Thus, we can say that the features of the realized state are a subset of the features of the realizing state.

[94] Another forerunner of the subset account can be found in Peacocke (1979). Peacocke defines the realization relation as "the indiscernibility relation with respect to causal contexts of a sentential kind" (117). Peacocke claims that x's being P is realized by y's being Q if and only if: every fact that x's being P causes to be the case, y's being Q causes to be the case as well (and vice versa), and every fact that causes x to be P also causes y to be Q (and vice versa). In effect, Peacocke requires realizers and realized properties to have indiscernible forward- and backward-looking causal features. This definition of realization is more restrictive than the subset account since the latter does not require that x's being P (the realized property instantiation) causes to be the case every fact that y's being Q (the realizer property instantiation) causes to be the case.

The account of realization, developed by Sydney Shoemaker (2001), Michael Watkins (2002), and Lenny Clapp (2001) explicitly uses the subset relation and focuses on a particular type of "feature": causal features. According to Shoemaker, a property X realizes a property Y just in case the forward-looking causal features bestowed by Y are a subset of the forward-looking causal features by X (and X is not a conjunctive property having Y as a conjunct).[95]

What are forward-looking causal features? They are the "contributions [the property's] instantiation can make to the causing of various effects" (Shoemaker 2003b, 2). Roughly, they are what Shoemaker used to call "conditional causal powers"; taken together, they are what determine a property's "potential for contributing to the causal powers of the things that have it" (Shoemaker 1980, 212). They are "conditional" because they do not specify the powers *simpliciter* that an object has when it possesses that property (i.e. what it can do in certain circumstances) but rather specify what powers a thing *would* have if it had *other* properties along with the property in question. To use a well-worn example, one of the causal features of *being knife-shaped* is *being able to cut butter if made of wood*; another is *being able to cut wood if made of steel*.

The ultimate subset realizer of any given property will be what Shoemaker (2003b) calls a "maximally determinate" property. This goes along with the fact that the subset realization is thought to capture the metaphysical aspects of the relation between determinates and determinables. For instance, *being colored* is subset realized by, say, *being blue*, which is in turn subset realized by *being cerulean*, which

[95] Instead of thinking of properties bestowing causal powers on their bearers, Clapp makes the "simplifying assumption" that properties are identical to sets of causal powers. With this assumption, Clapp defines realization as follows: "P realizes Q if and only if (def.), where p and q are the sets of powers constituting P and Q, q [is a subset of] p" (Clapp 2001, 129). See also Watkins (2002, 109, 115). Shoemaker (2003b) adds that the backward-looking causal features (e.g. being such that certain states of affairs are causally sufficient for the property's instantiation) of Y are a superset of X's backward-looking causal features.

is subset realized by *being a maximally determinate shade of cerulean*. Maximally determinate properties are those properties of a given kind (those that fall under a given determinable) which have maximal sets of causal features for properties of that kind. They have no distinct subset realizers because there is no property of that kind (had by the same entity) that has a larger, more encompassing set of causal features.

4.5.2.2. Micro-Realization and the Dimensioned Account

In his 2003b paper, Shoemaker discusses three other notions of realization in addition to the subset account (there called "realization$_1$"). The most important of these, realization$_3$ or *micro-realization*, is supposed to explain how a property instantiation is realized by a micro-physical state of affairs.[96] Realization$_3$ is designed to spell out the sense in which a series of (spatiotemporally and causally continuous) micro-physical states of affairs is a realizer of a career of an entity of a certain kind. The idea is that, in the most basic cases, an instantiation of a given maximally determinate property, *P*, (possessed by an entity with certain micro-entities as constituents) is realized in a micro-physical state of affairs type, *S*, just in case (1) whenever a token of *S* obtains, *P* is instantiated as well[97] and (2) the causal profile of *S* is isomorphic to that of *P* (Shoemaker 2003b, 13-16). (Where the *causal profile* of a property (or a type of states of affairs) is the set of all forward- and backward-looking causal features associated with the property (state of affairs type) – i.e. what an instantiation of the property (token of the state of affairs) can cause and what it can be caused by.)

Properties which are not maximally determinate can be realized$_3$ by microphysical states of affairs by being subset realized by maximally determinate properties. That is, a determinable property is realized$_3$ by a microphysical state of

[96] He also discusses two other kinds of realization, but these are derivative.
[97] And *vice versa* apparently, although this seems to rule out multiple realization.

affairs by being realized$_1$ by a maximally determinate property which is in turn realized$_3$ (as defined above) by that state of affairs (see Figure 1).

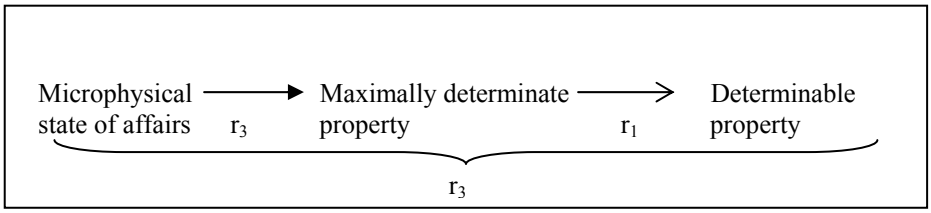

Figure 1: How Determinables Are Micro-realized by Microphysical States of Affairs
r_1 = subset realization
r_3 = micro-realization

Carl Gillett (2002) also has an account of the way in which a macroproperty of object *o* is realized in the properties and relations of the parts of *o*, which he calls the "dimensioned account." He introduces the dimensioned account by using it to characterize the sense in which the properties and relations that obtain between carbon atoms, such as being bonded and aligned in a certain way, realize the hardness of a diamond. According to the dimensioned account: "Property/relation instance(s) *F1-Fn* realize an instance of property *G*, in an individual *s*, *if and only if s* has powers that are individuative of an instance of *G* in virtue of the powers contributed by *F1-Fn* to *s* or *s*'s constituent(s), but not vice versa" (2002, 322). Gillett offers this as "a non-reductive, but [he] hope[s] illuminating, account of [realization's] connections to other basic notions" (ibid., 322 n.8), those other basic notions being, I presume, constitution, power, and property. As stated it is a disjunctive account. I will focus on the second disjunct, where the realizer properties are properties of parts of the entity that has the realized property, since this is the novel part of Gillett's proposal. The first disjunct, where the realizer properties are had by *s*, is merely a restatement of the subset account.

Micro-realization and dimensioned realization obviously disagree on the nature of the realizer. According to micro-realization the realizer is a state of affairs, a swarm of micro-entities bearing certain properties and relations to one another (the "concrete core"), together with positive and negative existential states of affairs that provide the context for the concrete core, the details of the concrete core's immediate environment. By contrast, according to the dimensioned account of realization the realizer is not a single state of affairs but rather a number of property or relation instances. In effect, the dimensioned account takes the realizer to be the properties and relations that make up a structural or "micro-based" property, taken severally.

4.6. Evaluating Causal Power Accounts of Natural Property Realization

Since I have argued that the explanation-based account collapses into some other form of natural property realization, we are left with the causal power accounts: the subset, dimensioned and micro-realization accounts.

First, consider the dimensioned account. It attempts to meet the location constraint by claiming that the realized property *P* contributes the causal powers it does in virtue of the fact that the realizer properties and relations of various micro-entities contribute the causal powers they do, and not vice versa. Whether or not this is satisfactory depends on what "in virtue of" amounts to in this context. It seems to me that "in virtue of," here, is just another way of picking out the realization relation itself. Thus, it is unclear if the dimensioned account of realization is an informative account of realization at all. The account is arguably viciously circular since it seems to appeal to the realization relation in what is supposedly a metaphysically informative account of realization. To be realized by something is just a special case of existing in virtue of that thing's existing, and Gillett's account tells us no more than this. The dimensioned account does not satisfy the location constraint; it is not clearly a

physicalistically acceptable relation between mental and restricted-sense physical properties.

I think that similar problems arise with respect to the causal process constraint. The dimensioned account cannot evade the problem of preemption that plagues the causal functional role account. According to Gillett, what distinguishes the dimensioned account from the causal functional role account (and the subset account) is that the dimensioned account does not take the notion of "causal role playing" literally (2002). This suggests that the dimensioned account is able to avoid identifying the realized property with the dimensioned realizer. This suggestion is supported by the obvious fact that the realizer is a number of distinct properties and relations which belong to entities other than the entity that has the realized property (e.g. the parts of the entity that has the realized property). However, without further details about the "in virtue of" relation, it would seem that it would lead to causal preemption of the realized properties by its realizers just as much as literal "causal role playing" does.[98] Thus, it is unclear whether dimensioned realization meets the causal process constraint.

Turning to the subset account, it is designed to satisfy the causal process constraint. At least for cases where the set of causal powers that characterizes the realized property is a *proper* subset of the set that characterizes the realizer, instantiations of the realized property will be related to instantiations of realizers as proper parts are related to wholes. So, the subset account vindicates the claim that realized properties are not identical to their realizers, since neither the proper subset nor the (proper) part-whole relation is identity. This also helps to defuse the exclusion problem since parts and wholes do not compete for causal sufficiency; so we can

[98] Cf. Beckermann (1992, 14) on the need to explain a non-causal use of "because of" or "in virtue of," especially if one does not appeal to property identities.

maintain the causal efficacy of realized properties.[99] The realized property and its realizer will both play a part in those causal processes that involve only powers in the subset that characterizes the realized property (which is why it is appropriate to describe these as broadly physical (e.g. mental) causal processes). However, only the realizer will be involved in causal processes that involve powers that do not belong to this subset.

How does the subset account fare with respect to the location constraint? Given that a subset realizer is restricted-sense physical, any properties it realizes obviously will be physical as well, since by definition the causal powers they bestow are a subset of the causal powers bestowed by the realizer. What about maximally determinate properties of macro-objects – can subset realization locate these in the restricted-sense physical domain? As discussed above these properties have no distinct subset realizers – they are self-realized in the subset sense. Yet they are not fundamental physical properties.[100] Further, not all of these maximally determinate properties of macro-objects are restricted-sense physical. So it seems that subset realization does not completely satisfy the location constraint – it seems to leave the restricted-sense physical basis of some properties unaccounted for. As Shoemaker puts it, the subset account does not address the question of "how the properties of fundamental particles and the like, and their relations to one another, realize the properties of entities composed of these" (Shoemaker 2003b, 6).[101]

---

[99] As we shall see in Chapter 5, this is related to the idea that subset realization is the metaphysical aspect of the relation between determinables and determinates.

[100] This is brought out by the inter-level exclusion problem – that the exclusion problem arguably generalizes so that properties of a macro-object are threatened to be causally preempted by the properties of and relations between the micro-entities that compose that macro-object.

[101] It turns out that microrealization is not more widely applicable than subset realization, at least in a physicalist world. (Thanks to Sydney Shoemaker for pointing this out.) That is, in a physicalist world, a property has a subset realizer if and only if it has a microrealizer. The left-to-right part of this biconditional is uncontroversial; all properties with subset realizers will have micro-realizers in a physicalist world. For, all entities must ultimately be composed of restricted-sense physical entities. What about the right-to-left conditional? Take some maximally determinate property that, as indicated above, apparently has only a microrealizer state of affairs of, say, type *T* and no distinct subset realizer.

Of course, micro-realization picks up exactly where the subset account leaves off. According to microrealization, maximally determinate properties are inclusive-sense physical because their causal profiles are isomorphic to (but need not be identical to) the causal profiles of microphysical states of affairs. Thus, it meets the location constraint, even with respect to maximally determinate special science properties.

Finally, how does microrealization fare with respect to the causal process constraint? Consider un-derived cases of micro-realization (unlike the derived cases illustrated in Figure 1). In these base cases, a maximally determinate property instantiation had by an object is micro-realized directly by a microphysical state of affairs. Is the maximally determinate property instantiation, *P*, both causally efficacious and distinct from the state of affairs, *Q*, that micro-realizes it? *P* is unquestionably causally efficacious, but it is not clear if it is distinct from *Q*. It might be argued that *Q* constitutes, but is not identical to, *P*. However, even if this is so and *P* and *Q* are not numerically identical, it is hard to see why *P* cannot be reduced to *Q*, if property instantiations and states of affairs are individuated causally. Recall that according to micro-realization the causal profiles of *P* and *Q* are *isomorphic*. Every feature in the causal profile of the subset realized property instantiation is preserved in the profile of the realizer. However, it is the proper subset relation that is the key to blocking reduction and satisfying part (b) of the causal process constraint. Only some features in the causal profile of the realized property are preserved in the realizer. This suggests that the lack of a formal analog of the subset relation in the account of

---

However, the entity that has the realized property will also have the property of *having a state of affairs of type T as part of its spatiotemporal career at the time it has the realized property in question*. This latter property will be a subset realizer of the maximally determinate property in question.
So, in a world in which physicalism is true, a property's having a subset realizer and its having a microphysical state of affairs type realizer are metaphysically equivalent. But microrealization is, in a way, explanatorily more powerful than subset realization. We need microrealization to provide a framework that explains why ultimate subset realizer properties are physical.

microrealization poses a problem for meeting the causal process constraint. There is no sense in which the instantiation of a micro-realized property is a *proper part* of its micro-realizer state of affairs.

In light of this, reexamining the argument presented above for the claim that subset realization satisfies the causal process constraint shows that it is not perfectly general. It depends on the causal features of the realized property being a *proper* subset of the causal features of the realizer. This is required if the properties are to be distinct, as demanded by part (b) of the causal process constraint. The subset account allows for cases where the improper subset relation holds between sets of causal features – cases of "self-realization." But in such cases it would be inconsistent to claim that there are two properties corresponding to a single set of causal features. It seems that the same will be true of all basic, un-derived cases of microrealization.

One might claim that this is not an untoward result. One might expect that maximally determinate properties are reducible to restricted-sense physical states of affairs types. However, while I think that some maximally determinate properties of macro-objects (e.g. a determinate density or mass) are reducible, maximally determinate *mental* and *biological* properties will also fall into the class that are so reducible, if this line of thought turns out to be correct. This result is unacceptable for any non-reductive physicalist.

The considerations presented here can be seen as posing a dilemma for the combination of subset realization and microrealization. Either maximally determinate properties are left simply self-realized by subset realization, which does not provide an acceptable restricted-sense physical grounding for them. Or via microrealization, one ends up reducing maximally determinate mental properties (and property instantiations) to restricted-sense physical states of affair types (and states of affairs).

That is, the combination of subset realization and microrealization appears to leave no room for non-reductive physicalism.

In next chapter, I show that this appearance is deceptive. By elaborating on some features that are implicit in the account of microrealization, particularly, some claims about the unnaturalness (from the restricted-sense physical point of view) of the microrealizers of mental properties, I argue that the dilemma outlined above can be avoided. However, reductive physicalists might wonder why one cannot define a realization relation whereby some *natural* microphysical property (or state of affairs type) realizes some mental property and only that mental property. I develop such an account of realization whereby natural microphysical properties (or states of affairs types) realize mental and other high-level properties. But, I argue that these natural structural realizers of mental properties will likely also simultaneously realize other high-level properties. If this is true, then there is no hope of reducing mental properties to natural restricted-sense physical ones. This will then set the stage for a reappraisal of the consensus view about what grounds irreducibility in Chapter 6.

CHAPTER 5

REALIZATION, THE DETERMINATE/DETERMINABLE RELATION, AND MECHANISMS

After discussing, in general terms, the metaphysical differences between the relation between mental and restricted-sense physical properties and that between determinables and determinates in Section 5.1, I argue in Section 5.2 that the way in which determinables are abstract relative to their determinates is different from the way in which mental properties are abstract relative to their physical realizers. I then introduce a notion of an aspect space for a kind of property in order to formulate this claim about abstraction more precisely. In Section 5.3, I discuss a phenomenon I call *multiple determinativity*, which, as I argue in the next chapter, is an overlooked possible basis for the irreducibility of mental properties. In Section 5.4.1, I outline two roles that mechanisms play in realization, sustaining and integrative mechanisms, and connect them to the claims about abstraction and multiple determinativity made in the previous sections. In Section 5.4.2, I develop an account of realization, *structural realization*, which gives a positive account of instances of subset realization that are not instances of the determinable/determinate relation and which may allow natural restricted-sense physical structural properties to realize mental and other special science properties. In Section 5.4.3, I show how to dissolve the apparent dilemma that I raised at the end of Chapter 4.

5.1. Two "Ways of Being" a Property – Two Kinds of Determination

My strategy in much of this chapter is to utilize the fact that some defenders of subset realization seem to hold that it is metaphysically indistinguishable from the determinate/determinable relation. I then argue that there are important metaphysical differences between uncontroversial cases of realization (e.g., pain/certain nerve fibers

firing in an appropriate sort of nervous system) and the determinable-determinate relation.[102] These metaphysical differences show that restricted-sense physical and mental properties are not related as determinates to determinables. So, the relation relevant to solving the problem of mental causation (and locating the mental in a physical world) is not the determinable/determinate relation. Thus, if the proponents of the subset account are correct and the determinable-determinate relation *just is* the subset realization relation, then subset realization is not an adequate account of the relevant relation between mental and physical properties.

However, in response, one may suggest (1) that subset realization is a more general relation, which has the determinate-determinable relation as a special case or (2) that subset realization and the determinate-determinable relation are each a special case of a more general realization relation, which is the one that is of interest in debates about physicalism and mental causation.

I accept that there is a broad or generic realization relation, characterized by the subset account, which can obtain in different ways. However, we lack a positive account of the cases of interest, cases of subset realization that are *not* cases of the determinable-determinate relation. I offer a positive account of such cases, what I call "structural realization," in Section 5.4. Further, since much of the utility of subset realization relation is derived from considerations regarding determinables and determinates (e.g. Yablo's proportionality constraints and their relevance to the exclusion problem), without saying more, it is unclear if these features will continue to hold for the cases of interest. Part of the significance of the next chapter is that it provides a framework to show that these features do hold for the mental/physical case.

[102] The conceptual machinery I develop to articulate these differences (if not the differences themselves) will also enter into my argument in the next chapter that multiple realizability is not the sole ground of irreducibility.

Proponents of the subset account of realization seem sympathetic to the idea that the relation between realized properties and their realizers is metaphysically the same as the relation between determinables and determinates (see, e.g. Shoemaker (2001), Clapp (2001), Watkins (2002)). They think that it is appropriate to talk of scarlet realizing red and of a neurological property of the brain being a determinate of a particular mental property. Stephen Yablo, the first to suggest the assimilation of these two relations, expected many to balk at it (1992, 256-7). He admits that there is a conceptual difference between the scarlet/red and physical/mental cases but challenges those who balk to find a metaphysical difference (ibid., 260).[103] Similarly, although Shoemaker admits that treating physical realizers as determinates of mental determinables "may depart from the traditional notion of the determinable-determinate relation" (2001, 85), it seems that this is *only* a conceptual difference (i.e. that there is no *a priori*, conceptual, or analytic connection between mental and physical terms – while there arguably is one between traditional determinates and determinables).

Yablo claims that "P determines Q iff: for a thing to be P is for it to be Q, not *simpliciter*, but in a specific way" (1992, 252). For instance, for a fire engine to be vermilion is for it to be red in a specific way, so *being vermilion* is a determinate of the determinable *being red*. Likewise, it seems right to claim that, say, *having C-fibers firing in a nervous system of the appropriate kind* is *having pain* in a specific way. So, according to Yablo's analysis *having C-fibers firing in a nervous system of the appropriate kind* is a determinate of the determinable *having pain*. However, I

[103] In a footnote, Yablo writes: "So *P* determines *Q* just in case the traditional relation's first, metaphysical component is in place, where this consists primarily in the fact that *P* necessitates *Q* asymmetrically. Probably it goes too far to identify determination with asymmetric necessitation outright; otherwise, for example, conjunctive properties determine their conjuncts and universally impossible properties are all-determining. For dialectical reasons, I try to remain as neutral as I can about where determination leaves off and 'mere' asymmetric necessitation begins" (1992, 253, n.23). I use paradigm cases of determinables and their determinates to argue that there are *metaphysical* differences between these relations.

think that the "specific way" is metaphysically different in these two cases. Consequently, these relations should not be assimilated to one another.

That there may be some metaphysical difference between the determinable/determinate relation and the relation between broadly physical properties and their restricted-sense physical realizers is suggested by the fact that there seems to be nothing to fill the blank in this analogy: to be in the physical condition K of this steaming tea is to be at 95°C in a certain *micromechanical* way just as to be the scarlet of Hester Prynne's letter A is to be red in a certain ___ way (see Yablo 1992, 253 for the tea example). While *being scarlet* is a way of *being red*, it makes no sense to try to characterize further the kind of mechanism by which scarlet, as opposed to crimson, determines red.[104] By contrast, not only is a physical realizer a way of being the mental property it realizes, it is perfectly coherent to ask *how*, or by what kind of mechanism, the realizer determines the mental property. Robbie the robot thinks that 2 + 2 = 4 in a mechanical way, I do so in a neurochemical way. Similarly, this paint stripper removes paint in a chemical way, that one does so in an electronic way; this braking system slows the car mechanically, that one does so hydraulically.

Providing a framework in which to spell out the physical manner in which mental properties are realized seems to be crucial to realization's being the relation that establishes that mental properties (and realized properties in general) are inclusive-sense physical. As I argue below in Section 5.4, specifying the manner in which a realizer is a way of being the realized property involves identifying a mechanism that is responsible for the instantiation of the realized property – what I will call a *sustaining* mechanism. Such a mechanism will involve the properties of

[104] If one adopted the view that colors are surface spectral reflectance properties, then this claim might not be true, depending on one's view of mechanisms. But the way in which determinables are multiply determined (if they are) is still different from one way in which realizers are multiply realized (if they are), as I discuss below.

and relations between microphysical entities that compose the object that has the realized property – what I will call the "components" of the restricted-sense physical realizer (perhaps along with relations to the environment). As I discuss below, physicalism does not demand that we specify or understand the workings of this mechanism (or how it results in the instantiation of the realized property). All that physicalism requires is that there be an adequate framework for elucidating this mechanism. Realization must provide this framework if it is to be able to underwrite explanations of how mental properties and powers are grounded in other physical properties and powers.

5.2. Abstraction and Property Aspects

The metaphysical difference adumbrated in Section 5.1 can be explicated by noting a feature of determinables/determinates discussed by W.E. Johnson and A.N. Prior. According to Johnson, there is no "secondary adjective" or differentia which can be added to *red* to distinguish *scarlet* from *vermilion* (1921, 176). As Prior puts it in his discussion of determinables: "We can say that the red and the blue agree in being coloured, but of their difference we can only say either that their colour is different or that one is red and the other blue" (1949, 5-6). Johnson's point is normally taken to be that a determinate is not a conjunction of a determinable and some other property. While this is true, I think it also suggests the point made above – that there is no sense in further characterizing the way of being red that scarlet (as opposed to vermilion) is, or the way of being colored that blue (as opposed to red) is. There is no way to characterize how, or in what way, scarlet's determining red mechanistically differs from crimson's determining red. We can state this point as follows: the only way that determinates of a single determinable differ is with respect to *aspects* of the determinable itself. For example, the only way in which (an instance of) scarlet differs from (an instance of) crimson is in its *redness*.

One plausible account of perceived colors sees them as individuated by three aspects: hue, saturation, and brightness. A determinate shade of *red*, like *scarlet*, is concrete in that it has a *particular* value of hue, saturation, and brightness. The determinable *red* is abstract in that it is less specific regarding these aspects; it is characterized by a *range* of hue, saturation, and brightness values. The very same aspects exhaustively characterize *both* the determinable and determinate.[105] In general, the hue, saturation and brightness values associated with a determinate color are a subset of those associated with its determinables. For example, the hue, saturation, and brightness values associated with the sequence – *vermilion*, *bright red*, *red*, *colored* – form a nested sequence of volumes of a *single* three-dimensional space, moving from *vermillion*'s very small, almost point-sized region to the entire space, which characterizes the ultimate determinable *colored*.

By contrast, two realizers of a mental property differ in ways that are not captured by aspects of the mental property itself. In order to specify how physical realizers of pain differ, we need not say that they differ in their "painness," in the aspects that characterize pains – for example, in the intensity, duration, affective or motivational aspects of the pain. Rather, we can list "non-pain" ways in which the realizers differ. For example, the realizers will be characterized by various aspects and distinguished along different dimensions, e.g., physiological, chemical, and physical. That is, the mental property is characterized by different aspects than those that characterize its realizer. It varies with respect to dimensions along which its physical realizers do not vary, and it need not vary along dimensions that characterize its physical realizers: qualitatively identical pains could be realized in different kinds of physical systems. Similarly, the aspects that characterize biological properties, like

[105] As I discuss below, this is related to the fact that a determinate falls under only a single hierarchy of determinables.

being a predator (e.g., whether it is specialized, generalized, apex, herbivore, or carnivore) or *being a gene* (e.g. whether it is dominant, recessive, structural, or regulatory) are different than those that characterize their restricted-sense physical realizers. In general, special science realized properties and their physical realizers differ from determinables and determinates in that they belong to different aspect spaces.[106]

This difference regarding aspects is reflected in the fact that mental properties are not abstract with respect to their realizers in the same way that determinables are abstract relative to their determinates. As noted above, a determinable is abstract relative to one of its determinates in that it takes up a larger volume of an aspect space common to both of them. A mental property is not abstract relative to one of its physical realizers in this way, but the way in which it *is* more abstract is harder to state precisely. If the intuitions behind the multiple realization argument are correct, mental properties are more "modally flexible" or "compositionally plastic" (Boyd 1980) than physical ones. They are paradigms of the properties Robert Stalnaker describes as "more abstract, and so might apply to things even if the properties on which they seem

---

[106] What exactly are these aspects of properties, like the hue, saturation and brightness of perceived color? I have no fully worked out, comprehensive theory to offer, but I will suggest a few possibilities. Perhaps aspects are qualitative or categorical second-order properties—properties of properties that partially account for their potential to enter into causal interactions. An instance of scarlet confers the powers it does in part because it has the hue, saturation and brightness it does. Thought of in this way, aspects are not sufficient to determine the causal features of a property, since relevant features of the environment (and perhaps contingent natural laws) are also required, but they are an important component of what determines a property's causal powers. (For some thoughts along these lines see Ellis (2005).) An alternative proposal is suggested by Worley (1997). Following David Armstrong, she writes of complex properties being analyzable into "constituent features" which are the "nature" or "intrinsic features" of properties. For instance, a triangle is a closed, three-sided plane figure; an equilateral triangle is a closed, three-sided plane figure with three equal sides. If one thinks of aspects as constituents in this sense, then one can model some cases of determination by claiming that the constituents of a *determinable* are a subset of those of its determinates. This obviously differs from the account sketched in the text, according to which determinables and determinate fall into a *common* aspect space and the values that characterize the determinate are a subset of those that characterize the determinable. On that approach, one could think of side length as an aspect of triangularity. *Being a triangle* leaves this aspect more unspecified than *being an equilateral triangle* does.

to supervene did not" (1996, 233). They could be instantiated over a wider range of physical conditions or situations than their realizers could.[107] Realized properties are abstract relative to their realizers in that they are relatively indifferent to which aspect space characterizes their realizers. For example, realizers of a mental property can be mechanical, neurochemical, hydraulic, etc. as long as these aspects "combine" or "aggregate" so as to result in the instantiation of the mental property.

Let's say that determinables are *homotopically* abstract relative to their determinates (since their abstractness is captured in the *same space*), while realized properties are *heterotopically* abstract relative to their physical realizers. Note that maximally determinate properties (specific shades of color) are just as heterotopically abstract relatives to their physical realizers as the determinables they fall under. They are just as general or modally flexible, just as indifferent to the aspects of the restricted-sense physical properties in virtue of which they are instantiated.

Here is a slightly different way to put the point. When we say that red is multiply realized we may mean two different things. We may mean (1a) that entities can have different "rednesses" – that *redness* is realized by *scarlet*, *vermillion*, and *titian*. Or we may mean (1b) that *redness*, even the same determinate shade of, say, vermillion, is realized by a variety of physical properties (and sustained by a variety of physical mechanisms): for example, by the way in which light is affected or produced by the properties and interactions of the particles in a bird's wing (diffraction), a dragonfly's wing (thin film interference), and neon and gas flames (incandescence).[108] Or consider the claim that temperature is multiply realized. We may mean (2a) that

[107] As I discuss in the next chapter, I think that multiple realizability is not what ultimately grounds the irreducibility and the abstractness of high-level properties. That is, despite what is suggested by the quotation from Stalnaker, a realized property will be abstract in this way even if it is uniquely realizable.

[108] This distinction would still be valid if we supposed that colors were identical to surface spectral reflectances.

things can have different temperatures (different determinate values of the same determinable). Or we may mean (2b) that a determinate temperature value can be realized in a variety of physical media by different properties and mechanisms: in gases by mean translational kinetic energy of the gas molecules, in plasmas (not in local thermodynamical equilibrium) by mean translational kinetic energy of the free electrons or by the mean translational kinetic energy of the ions, or in a vacuum by the blackbody distribution of transient radiation.

In cases (1a) and (2a), the realized property is homotopically abstract relative to its realizer, and both the realizers and realized properties are equally heterotopically abstract relative to the microphysical realizer which ultimately physically grounds each of them. In cases (1b) and (2b), the realized property need not be homotopically abstract (i.e. it could be a maximally determinate pain, color, or temperature value), but it is heterotopically abstract relative to its realizers.

This difference has been overlooked in the literature. For example, in Kim's recent discussion of realization, he does not appreciate that it is one thing to claim that *having a primary color* is realized by *being blue*, *being red* and *being green* and another to claim that *dormitivity* is realized by the chemical structures of *diazepam* and *secobarbital* (1998, 20-1). Kim is wrong to suggest that these two examples are exactly alike as instances of realization.[109] In general, a determinable bears two subset

[109] John Heil's (1999, 2003) deflationary treatment of multiple realizability fails for the same reason. Heil recommends that one adopt an ontology in which there are no determinable properties. He claims that "… many different things, many different kinds of thing, satisfy the predicate 'is red'. But do these things share a property? Are they identical (or exactly similar) in some one respect, a respect in virtue of which the predicate 'is red' holds true of them? Suppose, as I think likely, they do not. Some red things are crimson, some are scarlet, some rose, some magenta" (Heil 1999, 200). "Multiple realizability is not, as it is standardly thought to be, a relation among properties. It is simply a fancy name for the familiar phenomenon of predicates applying to objects in virtue of the possession by those objects of distinct, although pertinently similar, properties. This is characteristic of most of the descriptive predicates we deploy in everyday life and in the pursuit of science" (ibid., 203).

Note that Heil's response works only for a realization relation that is metaphysically indistinguishable from the determinable/determinate relation. It only works, to the extent that it does, for cases like scarlet realizing red, where the realizers are "similar but less than perfectly similar," that is, where they fall under the same determinable. It does not apply to alleged cases of multiple

relations to one of its determinates: (i) the range of values of the aspects that characterize the determinate are a subset of those that characterize the determinable, and (ii) the causal powers contributed by the determinable are a subset of those contributed by the determinate. I claim that the former subset relation does not hold for high-level properties and their microphysical realizers. The values of the aspects that characterize the chemical structure of *diazepam* are not a subset of those that characterize *dormitivity*. Likewise, the values of the aspects of a mental property are not a subset of those of its physical structural realizers. Thus, these physical structural realizers are not determinates of the mental properties they realize.

5.3. Multiple Determinativity

In this section, I discuss another metaphysical distinction between the determinable-determinate relation and the relation between mental properties and their physical realizers that is related to the fact that determinables and determinates belong to the same aspect space (while mental properties and their relevant structural physical realizers do not).

Although a given determinate will realize several different determinable properties, all of these determinables will fall under the same ultimate determinable. For example, every first-order property of individuals that a determinate shade of *red-orange* determines will be some more or less determinate *color*. Returning to some of Johnson and Prior's observations will help to clarify this point and show why it does not hold of physical realizers and the properties they realize.

---

realization, where a *maximally determinate* pain sensation (shade of red, etc.) is realizable by distinct neurological (or silicon) structures. For instance, two qualitatively (and causally) *indiscernible* instances of pain could be realized by different neurological structures in different organisms. Here we have perfect similarity in pain aspects coupled with (at least prima facie) dissimilarity in chemical or restricted-sense physical aspects. If physical realizers are not determinates of mental determinables, then Heil's tactic of eliminating determinable properties does not amount to a deflationary treatment of the relevant (multiple) realization relation.

Johnson claims that color is "distinctly other" than determinables like shape or tone (1921, 174). "Further, what have been assumed to be determinables—e.g. colour, pitch, etc—are ultimately *different* in the important sense that they cannot be subsumed under some one higher determinable, with the result that they are incomparable with one another" (ibid., 175).

For any two determinates, *X* and *Y*, of a given ultimate determinable *D*, every property determined by *X* falls under the same ultimate determinable as every property determined by *Y*. No property determined by *X* falls under an ultimate determinable other than *D*. Two shades of red are comparable to each other and every other color, but are incomparable to shapes. No shade of red can realize a shape. Two determinate properties either fall under the same ultimate determinable or fall under different ultimate determinables, but they cannot do both.[110] Since *scarlet* and *crimson* are two subset realizers of *redness*, they cannot realize any determinable that does not fall under *color*. This is because properties that fall under different ultimate determinables are *incomparable*, as Johnson points out.

Now consider properties that fall under the same ultimate determinable. Johnson claims that "… the several colours are put into the same group and given the same name colour, *not* on the ground of any partial agreement, but on the ground of the special kind of difference which distinguishes one colour from another; whereas no such difference exists between a color and a shape" (1921, 176). As Prior notes, Johnson was struck by the *incompatibility* of the determinates that are all the same

[110] As Dick Boyd pointed out (personal communication), these claims will hold only for non-conjunctive properties, since the property of *being a red square* will of course determine both *being red* and *being square*. Further, he raised the example of "cross-modal" adjectives like *warm* and *grating*, which might be used to compare colors and tones, and thus might be thought of as determinables covering both colors and tones. One way to handle such cases is to note that such adjectives are properties of properties and restrict the requirements on determinables and determinates to properties of individuals. Further, if one adopted a physicalist theory of colors, one could claim that *being a particular shade of fluorescent yellow* does not by itself determine *being grating* but does so only when conjoined with some property of the human nervous system.

level under a given determinable – by the fact that, for example, nothing can be both red and blue at the same time (in the same part).[111] This "special kind of difference" or incompatibility can be nicely captured within an aspect space. What unites determinates of a determinable is simply the fact that they all have different locations in the same aspect space. Two shades of red are similar in that they are close together in hue-saturation-brightness space; they fall within a certain volume of that space, but they are different colors simply because they occupy different regions in that same volume – the very thing by which they are grouped together. The *incompatibility* of determinates under a common determinable is captured by the fact that any part of an object can only have one determinate property from a given aspect space at a time. It is metaphysically impossible for an ultimate determinate property to be located in two particular locations in that aspect space.

If physical properties are determinates of the mental properties they realize, then two claims follow about them. First, all properties determined by a realizer, *P*, of a determinable mental property, *M*, should be *incomparable* to determinables that fall under an ultimate determinable other than that which *M* falls under. That is, *P* should not realize any properties that fall under a different ultimate determinable than that under which *M* falls. Second, *P* should be *incompatible* with any other (maximally determinate) realizer of *M*: no two maximally determinate realizers of *M* are instantiable in the same place at the same time.[112] However, as I argue below, neither of these claims holds, which shows that mental properties are not determinables of their physical realizers.

---

[111] One might think that determinate sounds or tones are not incompatible in this way. However, while two sounds may be present in the same room at the same time, their waveforms will occupy different regions of that room at any given instant (or will interfere to produce a third tone).

[112] After first drafting this chapter, I discovered that Doug Ehring makes a similar point (a principle that he calls "exclusion") in his (1996). Ehring also argues that the relation between mental properties and their physical realizers is not a species of the determinable/determinate relation.

Suppose the physiological property *having Aδ-fibers firing in a nervous system of the appropriate kind* realizes *having an acute pain*.[113] As just noted, if *having Aδ-fibers firing in a nervous system of the appropriate kind* is a determinate of *having an acute pain*, then it should determine no properties that fall under a determinable other than *having pain*. Further, the same should be true of all the realizers of this physiological property, including one of its restricted-sense physical structural realizers, say, *S*, an immensely complicated structural property of the organism involving properties of and relations between molecules and ions.

However, *having Aδ-fibers firing in a nervous system of the appropriate kind* realizes *having myelinated fibers firing in a nervous system of the appropriate kind* which does not, in turn, realize *having pain* or even *having a sensation* (supposing that is the ultimate determinable under which *having pain* falls). The restricted-sense physical realizer *S* realizes an even greater variety of properties of the nervous system in question: *having a certain mass*, *having a certain charge, elasticity,* and *electrical conductivity*. These two realizers of *having an acute pain* simultaneously realize properties that fall under different ultimate determinables and thus are not incomparable to those other determinables.[114]

---

[113] Note that this is a total realizer – a realizer whose instantiation is metaphysically sufficient for (necessitates) the instantiation of *having an acute pain* – not a core realizer (such as *having Aδ-fibers firing*), whose instantiation does not necessitate the instantiation of the realized property.

[114] Both of these examples are instances of "token" multiple determinativity, in which a single realizer token simultaneously realizes several different kinds of property instantiations. Token multiple determinativity entails type multiple determinativity. If a structural realizer token simultaneously realizes different kinds of property instantiations, then there is some type (property or state of affairs type) under which that token falls that realizes different kinds of properties. However, it seems consistent for there to be type multiple determinativity without token multiple determinativity. That is, a total structural realizer type could realize different kinds of properties, but no token of that type simultaneously realizes different property instantiations. However, in order to be interesting, this phenomenon must be distinguished from cases where a single *core* realizer type realizes different properties in different systems. In cases of type-but-not-token multiple determinativity, the structural realizer is a *total* realizer because it is metaphysically sufficient for any of its realized properties at the time it is instantiated (given certain background conditions). Nothing is left out, at that instant and given certain conditions, for the instantiation of one of its realized properties. Which of its realized properties is instantiated on a given occasion is determined by background context and what has happened in the system and environment up to the time of instantiation (what might be thought of as

Likewise, two realizers of *having an acute pain* need not be incompatible; it is metaphysically possible for two realizers of *having an acute pain* to be instantiated at the same time in the same part of an object. For example, suppose that *having Aδ-fibers vibrating in a nervous system of the appropriate kind* also realizes *having an acute pain*. Then, in an organism with a redundant nervous system that works both electrochemically and mechanically, *having Aδ-fibers firing…* and *having Aδ-fibers vibrating…* will both realize *having an acute pain* in that organism. Since a single Aδ-fiber could both be firing and be vibrating at the same time, these realizers are not incompatible.

These considerations suggest that many property realizers are *multiply determinative*: they realize high-level properties that fall under different determinables.[115] This terminology is meant to contrast with cases where a realized property is multiply *determinable*, i.e. multiply realizable. Figure 2 utilizes another example of multiple determinativity: Putnam's (1974) discussion of a cubical peg. The structural property instantiated by the rigid lattice of atoms that make up the peg

---

different triggering or background conditions). This contrasts with the aforementioned cases of core realization, in which the core realizer is not metaphysically sufficient at any time for any of its realized properties. In any case, I will be concerned with token multiple determinativity in what follows.

[115] Several authors have discussed relations that are similar to multiple determinativity: Menzies (1988), Kincaid (1988), Gasper's "multiple supervenience" (1992), and Endicott's "constructival plasticity" (1994) and "multiple realization complement" or "context sensitivity" (1998). Menzies discusses many interesting examples where two properties supervene on the same base property; multiple determinativity is most like the examples he discusses under "typical causal role supervenience" (except it is explicitly about total realizers). Philip Gasper (1992) discusses a similar phenomenon, which he calls "multiple supervenience," and argues that it blocks reductive explanation. But Gasper cites an example of what Menzies calls "logical supervenience" where a hierarchy of determinables (being less than 1 cm in length, being less than 0.9 cm in length, etc.) supervene on a single molecular base (1992, 668) as an example of multiple supervenience. Determinates of a given determinable obviously exhibit this phenonemon, but I claim that these determinates are *not* multiply determinative (in the right way)—they do not realize different *kinds* of properties, properties which fall under different ultimate determinables. Menzies' and Gasper's discussions are too broad—conflating relations that should be kept distinct. Endicott's "constructival plasticity" is simply the phenomenon of a single core realizer realizing incompatible properties when embedded in different total realizers. By contrast, multiple determinativitiy is not the converse of multiple realizability. It is not the claim that a (core) physical realizer *could have* realized different mental properties from those it actually does. I suspect that the failure to distinguish realization from supervenience has obscured the existence and significance of multiple determinitivity.

realizes the specific mass, color, and rigidity of the peg, in addition to its shape. Similarly, the atomic structure of a metal (the core realizer of which is its free electrons) realizes the metal's ductility, conductivity, opacity, and luster (see Menzies 1988).[116]

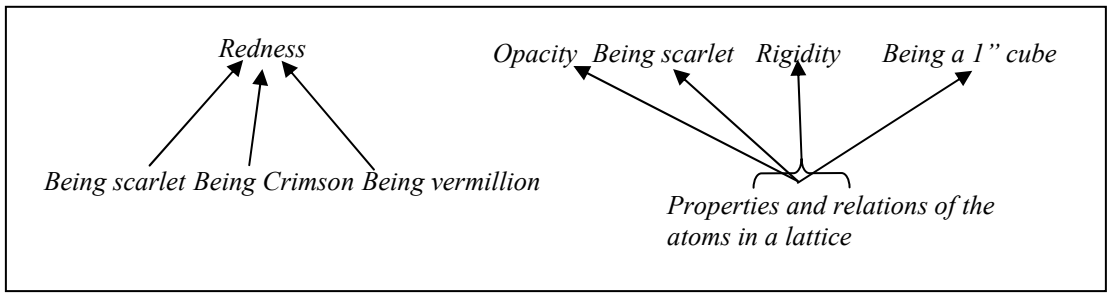

Figure 2: Multiply Determinable (Realizable) and Multiply Determinative Properties Compared
The figure on the left shows a paradigmatic multiply realizable property, redness, and some of its subset realizers. The figure on the right gives an example of a multiply determinative micro-structural realizer and some of the properties it realizes. "__→__" should be read as "__ determines __".

Note that multiple determinativity does not undermine the supervenience of the mental on the physical. Supervenience just requires that the distribution of the physical properties fixes the distributions of mental properties within and across worlds. It does not specify *how* this fixing is done. In particular, it does not say whether every physical property realizes a single high-level property or whether some

[116] Shoemaker mentions what amounts to an extreme version of multiple determinativity in connection with the subset account: "It might be supposed that if we start with the set of causal features of a self-realized property, there will be a property associated with every subset of this set, and each of these will have the self-realized property as a realizer. If this were so, then what is grounded in the self-realized property would not be a single hierarchy but a very complex treelike structure" (2001, 85). Shoemaker goes on to argue that this situation is not possible, and I agree. However, if the subset realizers are just determinates of determinables, then, as discussed above, there can only be a *single* nested hierarchy of realized properties (in accordance with the features of the determinable-determinate relation discussed by Johnson and Prior). But it is consistent with the subset account that there be multiple overlapping hierarchies of properties realized by a single self-realized property. That is, not *every* subset of the set, S, of causal features associated with a self-realized property corresponds to a property, but it may be that multiple subsets of S correspond to properties. So, self-realized properties (i.e. ultimate subset realizers) can be multiply determinative.

physical properties realize more than one high-level property. Multiple determinativity is not a modal phenomenon in a way that would violate the supervenience of the mental on the physical; it is not the claim that a physical property *could* realize other high-level properties than it actually does. So it is not the converse of multiple realizability; it is not a one-many relation between a set of base properties and a set of supervenient properties. It does not require that there be a mental difference between worlds without a physical difference between those worlds.

5.4. Mechanisms

5.4.1. Sustaining and Integrative Mechanisms

I suggested in Section 5.1 that the metaphysical differences between the determinable-determinate relation and the realization relation are connected to the fact that realizer types provide a mechanism for the instantiation of realized properties while determinate properties do not provide a mechanism for the instantiation of determinables. In this section I develop these claims about mechanisms to further clarify these metaphysical differences.

The following passage suggests a role that mechanisms might play in the realization of high-level properties:

> When P is said to 'realize' M in system s, P must specify a microstructural property of s that provides a causal mechanism for the implementation of M in s; moreover, in interesting cases – in fact, if we are to speak meaningfully of 'implementation' of *M – P* will be a member of a family of physical properties forming a network of nomologically connected microstructural states that provides a causal mechanism, in systems appropriately like *s*, for the nomological connections among a broad system of mental properties of which *M* is an element. These underlying microstates will form an explanatory basis for the higher properties and the nomic relations among them. (Kim 1993a, 343-4)

I think these points are related to the explanatory claims I made above in Sections 5.1 and 5.2. This mechanistic requirement is an essential part of the role realization plays

in locating mental properties in the physical world. Recall Fodor's discussion of the goals of reduction mentioned in Chapter 1:

> …the classical construal of the unity of science has really misconstrued the *goal* of scientific reduction. The point of reduction is *not* primarily to find some natural kind predicate of physics co-extensive with each natural kind predicate of a reduced science. It is, rather, to explicate the physical mechanisms whereby events conform to the laws of the special sciences. … [The two projects] are likely to come apart *in fact* wherever the physical mechanisms whereby events conform to a law of the special sciences are heterogeneous. (1974, 107, italics in original, underlining added)

Fodor is *not* arguing against the goal of reduction when it is construed mechanistically. Rather, he is arguing that this goal should not be confused with the traditional, logical empiricist construal of reduction.

Shoemaker provides what I think is an example of the mechanistically construed goal of reduction: a sketch of the microphysical mechanisms that are responsible for the chemical generalizations involving acids. If we suppose that being an acid can be realized by many different microstructural chemical properties, we expect there to be "something in common to the various realizer properties that *explains* their shared causal features" (Shoemaker 2001, 91, italics in original). Shoemaker notes that it is *not* the subset realization relation that is explanatory in this case.

> The explanatory relationship will be between, on the one hand, certain properties of, and certain relations between, protons and other subatomic particles and, on the other, the macroproperties of assemblages of these that have the conditional powers that go with acidity. The former will be the properties and relations that make something a proton donor and account for the fact that proton donors do the things acids do. Precisely what properties and relations of these sorts are involved will vary from acid to acid... (ibid.)

I think that Ned Block captures one way in which mechanisms figure in such explanations in his elaboration on some comments from Kim's (1992) paper:

> One common notion of realization … appeals to families of properties. The relations among temperature, pressure, entropy, etc are mirrored by relations among mean molecular kinetic energy, momentum exchange, etc, and the latter family provide[s] a *mechanism* for explaining the relations among the former. That is what makes the latter properties realize the former, or anyway it is closely connected to what makes for this realization. (Block 1997, 118, italics added)

Similarly, in the acid example the microstructure of acids explains "why proton donors do the things acids do" (dissolve metal, form a salt when combined with a base, turn litmus paper red, etc.) by being a member of a family of structural realizers whose causal relations mirror those between the properties of acids, metals, bases, litmus paper, etc. This holds because the family of realizers provides the causal mechanisms involved in the production of these acidic phenomena. This might be called the *integrative* role of mechanisms (because the mechanism illustrates how the realized property is integrated into relations with other realized properties). An integrative mechanism is a particular unfolding or development of the causal profile of a property. Here the realizer type *X* of property *Y* is part of a family *F* of realizer types which provides an integrative mechanism that can be used to model the causal relations between *Y* and the properties that are realized by other members of *F*. This is the goal of reduction at which Fodor gestures and which Kim and Block elaborate, according to which the physical realizers provide a mechanism that explains the relationships between the special science properties they realize.

However, I think that there is another, equally important, role that mechanisms play. This role is implicit in the claim that the relevant microstructure "makes something a proton donor" and hence an acid.[117] The realizer of a particular acid will not only be cited (along with other properties) in a description of the causal relations into which the acid enters but will also provide details about how the property of being

[117] Or as Kim puts it in the passage above, the realizer provides "an explanatory basis for the [instantiation of the] higher properties."

an acid itself is instantiated and how its instantiations persist through time.[118] This is what I call the *sustaining* role of mechanisms.

In order for a sustaining mechanism to be instantiated, the realizer *X* of a property *Y* must be a structural property or state of affairs type whose instantiation necessitates and *accounts for* the instantiation of the realized property. The structural realizer of *Y* will typically involve the properties of and relations between the parts of *O*, the entity that has *Y* (and perhaps parts of *O*'s environment) that are directly relevant to the instantiation of *Y*. Thus, the sustaining mechanism will involve the properties of and relations between these various parts of *O* – i.e. it will involve the components of the structural realizer. For example, *having brakes engaged* is realized by a structural property of my bicycle that consists of the properties of and relations between the brake levers, cables, wheel rims, and brake pads (inter alia), which provides a physical mechanism for the instantiation of the realized property. And *having brakes engaged* is also realized by a distinct structural property of a bicycle with a hydraulic braking mechanism. In each case, the realizer provides a mechanism that sustains the realized property throughout the period in which it is instantiated.

However, unlike the integrative mechanism, the sustaining mechanism is not the basis for a description of the causal relations between Y and *other* special science properties. It is not a mechanistic characterization of laws or generalizations that are the typical subject matter of the special sciences. Rather, it is a mechanism that explains how it is that *Y* is instantiated at a given time (and persists over time), which is often taken for granted by much of the research in the special sciences. Further, while integrative mechanisms are provided by a *family* of realizers, one of which is the

[118] Boyd (1999, 93-6) discusses this phenomenon and refers to it as "replacement stability." This is also what Fodor might have had in mind in the passage quoted above. When a property is multiply realizable, different instantiations of it will be explained by different sustaining mechanisms in different systems.

realizer of *Y*, sustaining mechanisms are provided solely by the components of the structural realizer of *Y* itself (and perhaps relations to the environment) – for instance, by the microphysical features in virtue of which a substance is a proton donor.

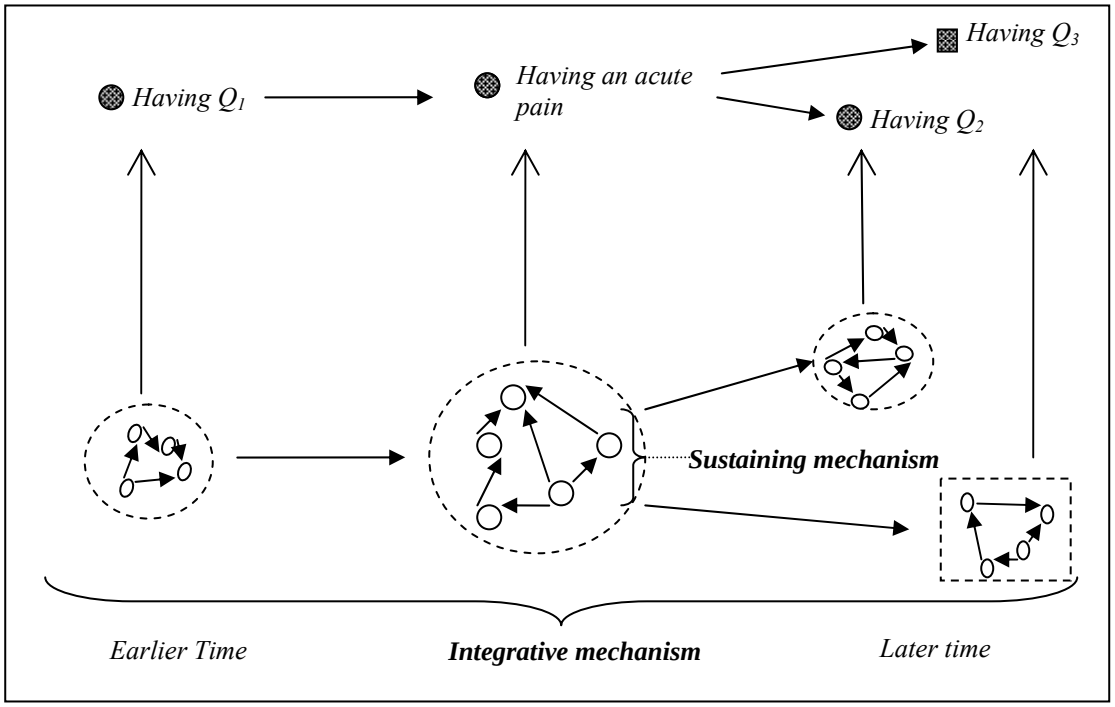

Figure 3: Schematic of Sustaining and Integrative Mechanisms

→ = *Causal relation*
⇢ = *Realization relation*
X = *Object instantiating property X*
⬚ = *Structural realizer token*

5.4.2. Mechanisms and Realization

This discussion of mechanisms meshes nicely with the metaphysical differences between the determinable/determinate relation and the realization relation between mental and natural restricted-sense physical properties. In this section I show how property aspects help spell out what it means for realizers to "provide" a sustaining mechanism for realized properties and how integrative mechanisms can be used to flesh out what multiple determinativity amounts to.

Recall that high-level properties are heterotopically abstract relative to their restricted-sense physical realizers; they belong to different aspect spaces, and the realized properties are indifferent regarding which aspects characterize their realizers. By contrast, determinables and determinates share the same aspect space. Hence, determinables are homotopically abstract relative to their determinates, and special science determinables and their determinates are both equally heterotopically abstract relative to their restricted-sense physical realizers. I think that these claims are connected to the fact that determinates do not provide sustaining mechanisms for the properties they determine.

Properties for which restricted-sense physical sustaining mechanisms are needed are biological and mental properties that belong to a macroscopic object but (physically) could not belong to that object's basic microphysical parts (like *being alive* and *being in pain*). Sustaining mechanisms give details about the properties of and relations between parts of an object that are responsible for the instantiation of that object's realized properties. This suggests the following requirement on sustaining mechanisms. In order for a realizer to provide a sustaining mechanism for the property it realizes the realizer must be characterized by different aspects than those that characterize the realized property. A sustaining mechanism would be uninformative and circular, in a sense, if it included the same type of property (a property characterized by the same aspects) whose instantiation the mechanism is intended to explain. This is why determinates of a given determinable do not provide sustaining mechanisms for the properties they realize; they belong to the same aspect space as their realized properties. This requirement (that the aspects that characterize the properties in the sustaining mechanism be different than those that characterize the realized property) also explains the popular image of mechanism's "underlying" the property whose instantiation it explains.

Of course, there are some properties of macro-objects which are physicalistically acceptable even if this requirement on mechanisms and aspect spaces does not hold. For example, macroproperties like *having mass* of *having a density greater than that of water* are subset realized by various determinates, such as *having a mass of 10kg* and *having a density of 5 g/ml* and the instantiations of these are microrealized by certain microphysical states of affairs. The fact that properties like *having mass* of *having a density greater than that of water* are subset realized is sufficient to establish that they are physical. I suggest that this is because these properties are what might be called molar properties of macro-objects.[119] Such properties contrast with what I will call "articulated" properties of macro-objects whose realizers must be very specific kinds of structures.

Articulated properties depend upon a complicated system of restricted-sense physical entities for their instantiation (in a physicalist world), while molar properties may be instantiated in any old amorphous blob of stuff. Molar properties do not require the existence and maintenance of a complex system or organism. Some other examples of molar properties are *having a radiodensity of 45 Hounsfield units* (as determined by a computerized tomography scan) and *being composed of 12% carbon*. By contrast, *being alive* and *being able to compute the square root of 2* are plausibly articulated properties. This distinction will be important in another context below (and in the next chapter).

The molar/articulated contrast seems to be relevant to the following distinction. Although no fundamental physical entity in fact has a mass of 10kg, it would be consistent with physicalism if one were to. I think this is true of molar macroproperties in general. By contrast, no fundamental physical entity has any

[119] I use the term 'molar' because such properties are concerned with unstructured, *masses of matter*, not a complicated microphysical structure.

articulated macroproperty, and if a property is such that it would be inconsistent with physicalism for a fundamental physical entity to possess it, then that property is articulated. For instance, it would not be consistent with physicalism for a photon to be alive, or feel pain, or be conscious, or metabolize. The molar/articulated distinction thus gets close to an elusive contrast between different kinds of "weakly emergent" property: between macro-properties that is would be physicalistically acceptable for fundamental physical entities to have and those that it would not be physical acceptable for them to have. All molar properties fall into the former class, and all properties that fall into the latter class are articulated.[120]

I think this difference is intimately related to the fact that the same type of mechanism that is responsible for the instantiation of a molar property in a macro-object is also responsible for the instantiation of properties of the same type in microentities. For instance, the same type of basic physical mechanism is responsible for the instantiation of *having a mass of 10kg* and for the instantiation of *having a mass of 2.99 x 10^{-26} kg* (the molecular mass of water) and the instantiation of *having a mass of 1.67 x 10^{-27} kg* (the rest mass of a proton). The possibility of a non-physical mechanism for the instantiation of molar properties is thus already ruled out. By contrast, the instantiation of articulated properties and the bestowal of their causal powers requires novel, complicated mechanisms that are not involved in the instantiation of any properties of more basic, microphysical entities. This opens up the possibility that these mechanisms are non-physical. In order for physicalism to be true, this possibility must be ruled out.

While property aspects and differences in abstractness are tied to sustaining mechanisms, multiple determinativity is connected to integrative mechanisms. The

[120] However, there may be some articulated properties that fall into the former class (i.e. which it would be physicalistically acceptable for some fundamental physical entities to have).

different kinds of properties that are simultaneously realized by a physical realizer correspond to different integrative mechanisms in which the structural realizer plays a role – different ways in which the causal profile of the realizer can unfold or develop over time. For example, the molecular structure of a metal will be a component in an integrative mechanism that describes the relations between its electrical conductance[121] and other macroscopic properties, but it will also be a component in a different integrative mechanism that describes the relations between rigidity and other macroscopic properties. Hence, a single sustaining mechanism type (given by the components of the structural realizer) can play a role in multiple types of integrative mechanisms.[122]

I propose the following positive account of the particular sort of realization relation that is not the relation between determinables and determinates, which I will call structural realization.[123] A restricted-sense physical structural property *X* structurally realizes property *Y* if and only if (a) the causal profile of *Y* is a subset of the causal profile of *X* and (b) *X* provides a sustaining mechanism for the instantiation of *Y*, which fits into a family of realizers that, in turn, provides an integrative mechanism for *Y* (a mechanism that models *Y*'s interactions and relations with other high-level properties). As discussed above, when property (or state of affairs type) *X* provides a sustaining mechanism for special science property *Y*, *X* must belong to a different aspect space than *Y*.

---

[121] Conductance (like resistance and current) is a macroscopic property of a sample of a substance as a whole, as opposed to conductivity (like resistivity and current density), which is a microscopic property, having values at every point in a body.

[122] In the next chapter, I discuss how different sustaining mechanisms can correspond to the same integrative mechanism, which amounts to substantive multiple realizability. So the relation between sustaining and integrative mechanisms is in general many-many.

[123] As noted in Chapter 4, we can freely translate between a state of affairs type *T* and a complicated property (*having a state of affairs of type T as part of its spatiotemporal career at the time it has the realized property in question*). A structural realizer will be one of these properties or states of affairs types. For simplicity, I will use properties.

I think that structural realization is a variety of realization whereby *natural* restricted-sense physical structural properties realize mental and other high-level properties. However, this still poses no threat to non-reductive physicalism, as I argue in the next chapter. In the final section of this chapter, I turn to dissolving the apparent dilemma that I raised at the end of Chapter 4.

5.4.3. Microrealization and Non-reductive Physicalism

The claims about mechanisms discussed above, and corresponding claims about natural properties which are implicit in the account of microrealization, can be used to show that the combination of microrealization and subset realization does not in fact require the reduction of special science properties (on pain of leaving those special science properties ungrounded). According to the account of microrealization, laid out in the last chapter, the causal profile of a given maximally determinate property instantiation, like *having an acute pain*, is isomorphic to the causal profile of a microphysical state of affairs type *T* which microrealizes it. The apparent problem I raised at the end of Chapter 4 was that the maximally determinate, ultimate subset realizer looked to be reducible to this restricted-sense physical state of affairs type. However, I think that this kind of "reduction" is no concession to the reductive physicalist.

While I agree that *T* is a state of affairs type that fulfills the requirements of microrealization, I contend that it is a natural mental type, not a natural restricted-sense physical type because it groups microphysical states of affairs only with respect to their mental causal features. The microphysical states of affairs of this type have *only* causal features that match up (via the isomorphism) with mental causal features of the realized mental property.

In other words, microphysical states of affairs of type *T* need not be unified in a way that is describable in terms of fundamental physical properties. Any token of

the state of affairs type *T* can determine *any* value of mass or charge and be made of any kind of material (e.g. myelin and acetylcholine, or silicon) as long as the causal contributions of each of these token realizers is the same relative to the mental property in question (i.e. as long as each realizes pain by means of the same process or mechanism). There are no constraints on the properties and relations that make up the microrealizers of this type, other than that they contribute to the instantiation of the realized property in the same way. In particular, there are no *fundamental physical* constraints on the components of the microrealizer. In other words, the instantiation of *having an acute pain* is indifferent to the aspects that characterize its microrealizers. *Having an acute pain* is heterotopically abstract relative to its microrealizer.

Finding a type that has a causal profile isomorphic to any given mental property depends on the world already being cut up into determinable/determinate hierarchies – being divided into natural kinds of mental properties. But the natural restricted-sense physical types under which complicated, microphysical states of affairs fall cut across these hierarchies. This is because these restricted-sense physical kinds are not characterized by the same aspects that characterize mental kinds. Only the microrealizers of *molar* macroproperties cleanly line up with determinable/determinate hierarchies of properties of macroscopic objects because these realizers and realized properties share a common aspect space. In short, microrealization allows for non-reductive physicalism because, given a determinate special science property, the microphysical states of affairs types that microrealize it *and only* it are not natural restricted-sense physical types.

However, the reductive physicalist might turn to structural realizers of mental properties, as defined in section 5.4.2. These will arguably be natural restricted-sense physical properties because they correspond to restricted-sense physical sustaining mechanisms (sustaining mechanisms involving only restricted-sense physical

properties and relations). However, I think that the empirical evidence suggests that any natural restricted-sense physical property that structurally realizes some mental property will be multiply determinative. As I argue in the next chapter, if the structural realizers of mental properties are multiply determinative in this way, this grounds the irreducibility of those mental properties. Thus, multiple realizability is not the sole ground of irreducibility. I also use multiple determinativity to address the explanatory exclusion problem (i.e. why the special sciences are autonomous) introduced toward the end of Chapter 3.

CHAPTER 6

MULTIPLE REALIZABILITY, MULTIPLE DETERMINATIVITY, AND IRREDUCIBILITY

There is overwhelming consensus in the literature that the question of whether a realized mental property is irreducible hinges on whether that property is multiply realizable. As Ned Block remarks, "For nearly thirty years, there has been a consensus ... that reductionism is a mistake and that there are autonomous special sciences. This consensus has been based on the argument from multiple realizability" (1997, 107). This opinion is shared by critics of the consensus view. For example, Kim writes:

> I expect many philosophers will reply that biological properties are no more physically reducible than psychological properties, citing their "multiple realizability" in relation to physicochemical properties. For most antireductionist philosophers, multiple realizability has long been their mantra, an all-purpose antireductionist argument applicable across the board to all special science properties. They see multiple realization everywhere, and this leads them to see irreducibility everywhere. (Kim 2003, 166)[124]

Recently, some philosophers have used the assumed connection between multiple realizability and irreducibility to argue that the consensus view of the special sciences needs to be revised. These arguments, which I discuss in detail below, use negative conclusions about the extent or significance of multiple realizability to draw negative conclusions about the irreducibility of special science properties.

The main claim of this chapter is that the basis for the consensus is mistaken. Multiple realizability is not always the best strategy for defending irreducibility and

[124] To cite just two more examples: Elliott Sober, in a discussion of what he calls Putnam's "curious remark" which suggests that multiple realizability is *not* in fact required for the autonomy of higher level explanations, also notes that most philosophers have taken multiple realizability to be essential to the "Putnam/Fodor argument" (1999, 549 n.8). (I discuss this "curious remark" in Section 6.6.) And Louise Antony thinks that "the most promising strategy for defending the reality of mental properties" is to "appeal to their multiple realizability" (1999b, 3).

the autonomy of the special sciences. In fact, multiple realizability is not even required for irreducibility. Rather, the irreducibility of higher level properties and special sciences is grounded in many-many relations between mechanisms, of which multiple determinitivity and substantive multiple realizability are different facets.[125] However, while my view challenges the basis for the consensus view, I think that it explains some passages from two of the originators of this consensus, which would otherwise be puzzling. More importantly, it also provides a metaphysical basis for non-reductionist claims about explanatory and theoretical autonomy that are challenged by the explanatory exclusion argument introduced toward the end of Chapter 3. (Recall that, in a nutshell, the explanatory exclusion argument concedes that mental and restricted-sense physical properties do not compete with respect to causal sufficiency, but claims that this is only because there are no natural mental properties. There are only restricted-sense physical causal processes, and the special sciences are just heuristic crutches that we humans rely on because such microphysical processes are relatively epistemically inaccessible.)

6.1. The Multiple Realizability Argument against Reducibility

The basic idea behind the multiple realizability argument is simple, indeed so simple that the argument is often not explicitly formulated. If mental properties were reducible to physical properties, then the relation between mental and physical properties would have to be one-one. But, so the argument continues, there is good reason to think that many mental properties are, or could be, realized by different

[125] As I discuss below, there are different kinds of multiple realizability. So the question of whether multiple realizability is *sufficient* for irreducibility depends on what should be required for *substantive* multiple realizability. What most philosophers have thought of as multiple realizability (simply a one-many relation between two sets of properties) is not sufficient for irreduciblity; it is merely a symptom or indicator of the stronger claim that there is a one-many relation between high-level integrative and restricted-sense sustaining mechanisms. However, if this claim about mechanisms is built into the claim that a property is multiply realizable, then multiple realizability is sufficient for irreducibility, as I discuss below.

physical properties in different organisms. So the relation between mental and physical properties is one-many. Hence, mental properties are not reducible to physical properties. In the most influential of Putnam's papers, (1967b, 436) and (1975b, 293, 296), multiple realizability is presented as an a posteriori argument, as it is in Fodor's "Special Sciences" (1974). However, in earlier papers (1960, 371) and (1964, 392), Putnam seems to present the multiple realizability of mental states as something that can be known a priori—as arising from the "logic" of psychological theories or from the definition of Turing machine states.[126] Further, the discussion in (1960), (1964), and (1967b) focuses on the irreducibility of mental states, while (1975b) focuses on the irreducibility or autonomy of higher-level explanations. An argument drawn from the former three papers can be formulated as follows:

1. If mental property *M* is reducible, then there is some natural, restricted-sense physical property *P* such that necessarily,[127] for all *x*, *Mx* iff *Px*. ("General" or "uniform"[128] property reduction is a necessary condition for reducibility.)

2. But for any natural, restricted-sense physical property *P* that necessitates (realizes) *M*, it is possible for some individual to lack *P* but have *M*, since some other natural, restricted-sense physical property, *Q*, could realize *M*. (*M* is multiply realizable.)[129]

3. The disjunctive property, $P \vee Q$, is not a natural, restricted-sense physical property.

---

[126] Putnam cites some of Fodor's earlier work in support of the a priori view: "Thus, as Jerry Fodor has remarked ... , it is part of the 'logic' of psychological theories that (physically) different structures may obey (or be 'models' of) the same psychological theory" (Putnam 1964, 392).

[127] I officially leave open whether nomological necessity or metaphysical necessity is operative. Many philosophers think that metaphysical necessity is too strong, since it rules out the existence of worlds that are physically unlike but mentally like the actual world. To cover as many positions as possible, I only rely on nomological possibility below.

[128] This is meant to contrast with "local" property identities, which are discussed below.

[129] The earlier 1960/1964 version of the argument and the latter 1967b version differ regarding whether (2) can be known a priori. There is a related difference regarding whether M has to be multiply realized for reduction to fail, or if it is sufficient for M to be merely multiply realizable. This question is tangential to my interest in this chapter, which is the general structure of multiple realizability arguments and criticisms of them. However, I think that are good reasons to go with the a posteriori version of the argument. For one thing, the a priori version seems to confuse properties and concepts (or Turing machines with machine tables, i.e. representations of Turing machines). Three classic a posteriori arguments for multiple realizability are found in Block and Fodor (1972).

4. So, there is no natural, restricted-sense physical property *P* such that necessarily, *Mx* iff *Px*.

5. So, *M* is irreducible.

Note that the domain of quantification in premise (2) is restricted. It does not claim that that for any physical property *P*, it is possible for something to have *M* but lack *P*. This is too strong as Yablo (1992, 255) points out.[130] Rather, we need to limit quantification to physical properties that realize *M*.

6.2. Critiques of the Multiple Realizability Argument

Several philosophers have claimed that the argument for property irreducibility fails. Indeed, each of premises (1), (2) and (3) has been challenged, respectively, by the local reduction strategy, what I will call "direct arguments against multiple realizability", and the disjunctive strategy. These arguments are sometimes used in tandem or entwined together, but I will discuss them separately as much as possible.

2.1. The Disjunctive Strategy

The disjunctive strategy challenges premise (3) in the multiple realization argument. In effect it claims that proponents of the multiple realizability argument have simply focused on the wrong physical properties. There is a physical property that realizes *M*, namely, the disjunction of all of the possible non-disjunctive total realizers of *M*, which I will denote by "$\vee P_i$", for which it holds that necessarily, *Mx* iff $\vee P_i x$. Thus, *M* can be reduced to this disjunctive property, $\vee P_i$. Of course, the burden of someone proposing this line of argument is to show that disjunctive properties exist—something that has been questioned by many philosophers.

Although the disjunctive strategy is sometimes attributed to Kim (e.g., as Antony (1999a) does), it is not clear that he ever fully endorsed it as a response to the

[130] I follow Yablo in holding that every mental property is multiply realizable if and only if: "Necessarily, for every mental property *M*, and every physical property *P* which necessitates *M*, possibly something possesses *M* but not *P*" (1992, 255).

multiple realizability argument. He seems to be ambivalent about the existence of disjunctive properties, and his position has changed over time.[131]

However, Lenny Clapp (2001) has recently offered a thorough defense of the disjunctive strategy. Consequently, he has suggested significantly weakening non-reductive physicalism – claiming that it is merely an "epistemic" or "pragmatic" thesis, due to our cognitive limitations, not to the structure of the world. Clapp formulates a dilemma based on some of Kim's work: either disjunctive predicates designate properties or they do not. If they do, then reductionism succeeds, since a purportedly multiply realizable property, *M*, will be at least nomologically coextensive with the property that is designated by the "properly disjunctive" predicate whose disjuncts are designators for each of the realizers of *M*. If no disjunctive predicate designates a property, then mental predicates do not designate "legitimate," projectible, natural properties either (2001, 120-1). For, in that case, mental properties would inherit the causal heterogeneity from the heterogeneous disjunction with which they are at least nomologically coextensive. Clapp thinks that the latter horn leads to "rampant illegitimacy" because almost all properties are multiply realized (ibid., 121-122).[132] So, he argues that some disjunctive predicates can designate genuine properties. And he argues that in light of this we should adopt a modified, weakened form of non-reductive physicalism.

---

[131] For example, Kim sometimes suggests that the disjunctive strategy may be viable: "These species-dependent correlations do not of course warrant the species-independent blanket identification of pains with a "single" brain state (assuming that we refrain from making up "disjunctive states," although I think this is an arguable point)." (1980, 235). Later, Kim notes that there are reasons for not countenancing disjunctive types (see his 1992) but goes on to sketch the following proposal: "any given M-instance must be either a P1-instance or P2-instance or …, where P1, P2, .. are realizers of M, and the set of all M-instances is the union of all these Pi-instances. In this sense, we may say that mental kind M is *disjunctively identified* with physical kinds P1, P2, … Note that M is not identified with the *disjunction* of P1, P2, ..; nor is an M-instance identified with an instance of the disjunctive property P1 or P2 or … We may call this proposal 'multiple-type physicalism' (1993b, 364, italics in original).

[132] I think it's false that almost all properties are multiply realizable (at least in a substantive way that has implications for reduction), as I discuss below.

Clapp argues that a "properly disjunctive predicate"[133] $\pi = (\pi_1 \vee \pi_2 \vee \ldots \vee \pi_n)$ designates a property *P* if and only if there is some nonempty set of causal powers *p* such that (a) if a particular *o* satisfies π then *o* possesses every power in *p*, and the converse, (b) if a particular *o* possesses every power in *p*, then *o* satisfies π (2001, 127)—for short, a property disjunctive predicate designates a property if and only if the disjuncts "satisfactorily overlap" on a set of causal powers (ibid., 132). He then argues that, assuming what amounts to the subset account of realization, the predicates that designate each of the possible realizers of mental property *M* will appropriately overlap on the intersection of the sets of causal powers associated with each of the possible realizers. Thus, the properly disjunctive predicate formed from predicates that stand for these realizers will designate a disjunctive property, which is at least nomologically coextensive with *M*.

Clapp works out in detail how this procedure works for the determinates of a given determinable. For example, the disjunction formed from all of the determinate shades of color will be a legitimate disjunctive property; the disjuncts will overlap on the intersection of the sets of causal powers associated with each determinate color. The set of causal powers in the intersection constitutes, according to Clapp, the multiply realizable, determinable property *being colored*.

This account of disjunctive properties allows the reductive physicalist to respond to the multiple realizability argument as follows. Although proponents of that argument can rightly claim that *M* is not reducible to any of the properties designated by the individual π_i's, they cannot claim that *M* is irreducible. In order to conclude this, one needs the additional assumption that there is no other physical property to which *M* can be reduced. The construction and legitimacy of the disjunctive property

[133] A "properly disjunctive predicate" is defined as follows: "a disjunctive predicate $(\pi_1 \vee \pi_2 \vee \ldots \vee \pi_n)$ is a properly disjunctive predicate if and only if (i) there is more than one disjunct π_i; (ii) each disjunct π_i designates a legitimate property; and (iii) each π_i designates a distinct property" (Clapp 2001, 123).

formed out of the individual, non-disjunctive realizers of *M* shows that this assumption is false.

Elliott Sober has also raised doubts about the assumption that all disjunctive properties are not natural. Sober notes that laws that "specify a quantitative threshold for some effect" seem to be disjunctive—for example, the law that water boils at sea level when its temperature exceeds 100°C (Sober 1999, 553). For this law simply claims that water boils at sea level when its temperature is 101°C or 102°C or 103°C... We just utilize a "handy shorthand for summarizing these disjuncts" – that any temperature above 100°C will produce boiling water. But, as Clapp argues, this just shows that "above 100°C" and a properly disjunctive predicate can designate the same (or at least a nomologically coextensive) property. Hence, some properties designated by properly disjunctive predicates can figure in laws. These properties are natural and are not gruesome or "wildly heterogeneous."

6.2.2. The Local Reduction Strategy

The local reduction strategy challenges the soundness of the multiple realizability argument by attacking (1). Put in schematic form, the local reduction strategy claims that reduction does not require that necessarily, for all x, Mx iff Px; rather, all it requires is that necessarily, for all x that are S_1, Mx iff P_1x, and necessarily, for all x that are S_2, Mx iff P_2x, ... , where the S_i are particular types of species or structures.

There are two slightly different versions of this strategy: what I call the eliminative version and the conservative version. The eliminative version concedes that general mental properties (those that could be said to apply to many different species of organism) may be multiply realizable and hence irreducible. However, such a general, uniform reduction is not needed because these general mental

properties are not natural.[134] Further, the local reductionist argues that, perhaps due to common evolutionary descent, there is likely to be a single physical property that realizes, say, *pain-in-S* for a given species, *S*.[135] Thus, we can identify *pain-in-octupi* with one physical property and *pain-in-humans* with another. According to this version, *M* is not a natural property and is "broken up" or "partitioned" into natural properties that are unique to particular species or structural types: *M-in-*S_1, *M-in-*S_2, etc. Then each of these species-specific high-level properties can be reduced to (identified with) the unique physical property, P_1, P_2, ... that realizes it in the respective species. I call this the eliminative version of local reduction, since there is no natural property, being in pain, that is common to all individuals that experience pain. Rather, there is just a concept of pain (or a property in the abundant sense) that is not a natural property and not causally efficacious. All that is there is to reducing *M* is reducing *M-in-*S_1, *M-in-*S_2, etc. because there simply is no natural property of being in *M* simpliciter. This version captures the local reduction strategy as presented in David Lewis' and (most of) Jaegwon Kim's work: see Lewis (1980, 233; 1994, 304-8) and Kim (1980, 235; 1992, 332-3; 1998, 109-111).

The second version of the local reduction strategy sees it as part of a package with the disjunctive strategy. This is a "conservative" version of the strategy since it maintains that there is a natural property of, say, *being in pain*; it conserves this population-independent property as a disjunctive restricted-sense physical property. According to this line, each of the species-specific higher-level properties is still

[134] Kim argues for the unnaturalness of mental properties using ideas that are related to the disjunctive strategy. He starts with the claim that disjunctive properties are not projectible; they are "wildly heterogeneous." Hence, the realized properties which are nomologically coextensive to these disjunctive properties are equally unprojectible (Kim 1992).

[135] One common anti-reductionist response to the local reduction strategy is to claim that a given mental property may be multiply realized within a species or even within an individual over time (Horgan 1993). To succeed, this response would have to show that these are instances of substantive multiple realization, discussed below in Section 6.2.3.1.

identified with the corresponding physical realizer. However, when one asks about the alleged property that all of the individuals from these various species have in common, this interpretation does *not* claim that no such natural property exists. Rather, it claims that this general property is reducible to the disjunctive property formed from all of these realizers. Thus, this strategy can only be used in tandem with the disjunctive strategy, since the only way to conserve this general property in the reductive framework is to identify it with the disjunction of its realizers (see Kim 1992, 332; 1999, 16).[136]

6.2.3. Direct Arguments against Multiple Realizability

Finally, there are arguments that move directly against (2) by claiming that the constraints on substantive multiple realizability are more restrictive than previously thought. By appealing to these constraints, these arguments claim that many purported examples of substantive multiple realization are merely trivial; and trivial multiple realization is not sufficient for irreducibility. Reduction can go through in these cases where the realizers have all their "significant" or "relevant" properties in common. In contrast to the local reduction strategy, which accepts the (largely intuitive) account of multiple realizability put forward by non-reductionists, according to which the individuation of realizers is fairly fine-grained, the direct arguments re-examine what is required for substantive multiple realizability. They claim that the requirements for a property being substantively multiply realizable are more stringent than non-reductionists imagined. Some of these arguments proceed by raising

[136] Kim (1992, 316, 332; 1998, 93, 106-110) offers the disjunctive strategy as one possible response to Ned Block's queries about local reductions. Namely, what do the pains of dogs, people, etc. have in common in virtue of which they are all pains? Is pain a "general" property distinct from each of its species-specific realizers? In response to these questions Kim offers the following suggestion: "… in the aftermath of local reduction, identify pain with the disjunction of its realization bases…". But Kim suggests that this option is not as plausible or satisfying as simply eliminating the general property of pain from one's ontology and taking the eliminative route.

questions about the individuation of realizers: for example, by asking why they cannot be individuated coarsely (or functionally) as realized properties are.

6.2.3.1. Substantive and Trivial Multiple Realizability

The direct argument depends upon a distinction between substantive and trivial multiple realization. Several philosophers have hit upon this distinction, apparently independently. However, these precedents in the literature have not been cited by proponents of the direct argument. Further, many authors have overlooked (or rejected) this distinction, and consequently have been led to see multiple realization as a nearly ubiquitous phenomenon (as Kim alleges in the passage quoted above). Because of the confused state of the literature, I think it is worth quoting extensively both from those who have made the distinction and those who have not.

The distinction between substantive and trivial multiple realization is implicit in David Lewis's admonition to "beware … of finding spurious variation by overlooking common descriptions" (Lewis 1994, 305). He provides an example of two mechanical calculators that are alike in design. When addition is performed, amounts that are carried go into a register that is selected by throwing a switch, but the location of the registers is different in each calculator. Lewis claims that we should not say that the carry-seventeen role is occupied in one machine by state of register A and in other machine by state of register B. Rather, we should say that the role is occupied in both machines by a state of the register selected by the switch. He goes on to suggest that a kind of thinking that some humans do on left side of brain and others do on right side may be analogous case. The point is that such cases should not be described as instances of multiple realization – to do so would trivialize multiple realizability. The realizers in these cases are not different in any significant way; the spurious variation in realizers is merely an artifact of arbitrary, overly fine individuation criteria. Taken to the extreme, this would conflate multiple realization

with multiple instantiation – where every object in which a realized property is instantiated is taken to have a unique property corresponding to a distinct "realizer." Ruth Millikan makes these points in the following suggestive passage:

> … is liver function multiply realized in humans because some people's livers are larger than others and have more cells? Are verbal abilities multiply realized if the neural structures responsible for them, though operating in exactly the same manner, occur more bilaterally in some people than others? In a well known passage, Putnam (1975) analogized the multiple realization of functional properties to the multiple configurations of colliding atoms that might cause a square peg to refuse to go through a round hole. Did he actually intend this as an *example* of multiple realization of a functional property? Then, it would seem, having a low center of gravity would be a multiply realizable functional property too? Jackson and Petit (1990) analogize existentially generalized properties, such as *some of its atoms are decaying*, to function[al] properties – because any of various individual atoms might be the ones that are decaying. Are Newton's laws also multiply realized because sometimes it is atoms of gold and sometimes of lead that make up the masses to which they apply? Or because sometimes the atoms are arrange in crystalline structure and sometimes not? Or because it might be the atoms named Sally and Mike that help make up the mass or it might be the atoms named Betty and Michael?
>
> Clarification is surely needed in this area, nor [sic] will I attempt much of it here. But something like the following distinction seems to be required. Sometimes different [token] mechanisms that accomplish the same [task] operate in accordance with different principles; other times they represent merely different embodiments of the same principles. Or we might say, sometimes looking more closely at the mechanism helps to explain *how* it works; sometimes it reveals only what stuff it is made of. It is only the former kind of difference that makes interesting "multiple realizability." (Millikan 1999, 61-2, italics in original)

According to the view suggested by this passage, holding that Newton's laws are multiply realizable and hence irreducible, in the same sense that mental properties allegedly are, would be untenable. They are not multiply realizable because different mechanisms are not involved in their different "realizations." Adopting such a view would result in the multiple realization argument overgeneralizing and becoming dialectically useless. If virtually every property is multiple realizable, then the

proponent of the multiple realizability argument is forced make the absurd claim that virtually every property is irreducible. One would thus fail to capture what is allegedly special about high-level mental, biological, and social properties.[137]

However, several philosophers seem to have adopted a view of multiple realization according to which it ubiquitous (or at least have been tempted by such a view). Here are a few examples:

> States like masses, volumes, and temperatures are even more variably realized than mental states: one can have a gram or a litre of almost anything, at any one of an indenumerable infinity of temperatures. (Crane and Mellor 1990, 197)

> It is in the nature of properties, however, that they are inevitably multiply realisable. If a given property, *P*, can be exemplified by objects of different sorts, objects possessing distinct properties, then *P* is, in one clear sense, multiple realisable. Regarded in this light, it is difficult to think of a property that is *not* multiply realisable. Consider *being green*. Many different sorts of things can be green: plants, inanimate objects, beams of light. This by itself does not show, however, that being green must be a 'functional property', or that being green is not type-identical with some perfectly ordinary physical property shared by plants, inanimate objects, and beams of light. (Heil 1992, 134)

> ... there are many properties that Kim would count as first-order physical properties that are multiply realizable, and for which the same apparent problem about projectibility arise. Consider the property of having mass of one gram. Just as in the case of jade, there will be some instantiations of this property that will possess causal powers different from those of other instantiations—and of course in the case of "having mass of one gram" the range of variation will be vastly greater. (Antony and Levine 1997, 91)

> … all, or almost all, properties are multiply realized. If the fact that different sorts of creatures can instantiate *being in pain* leads us to believe that *being in pain* is multiply realized, then the fact that many different sorts of things can instantiate, for example, *being green* ought

[137] As I discuss below, I believe that Millikan's distinction is related to the difference between molar and articulated macro-properties, which was introduced in Chapter 5. Looking at the details of the structural realizer of a molar property reveals only "what stuff it is made of," while looking at the details of the structural realizer of an articulated property reveals "how it works."

> to lead us to believe that *being green* is multiply realized. The point also holds for paradigmatic physical properties. Consider the paradigmatic physical property of having a mass of two grams. This property is instantiated by many different kinds of objects—bits of paper, bone, metal, jelly, and so on. … Claiming that a property *P* is *not* multiply realized is claiming that there are no properties in virtue of which objects instantiate *P*; it is tantamount to claiming that when an object instantiates *P* it is just a "brute fact" that it does so. (Clapp 2001, 121)[138]

I think these passages single out the kind of trivial multiple realizability that Lewis and Millikan warn against assimilating to substantive multiple realizability. The direct argument strategy develops criteria to distinguish trivial from substantive multiple realizability and argues that the way in which mental properties are multiply realized falls into the trivial category – one that has no anti-reductive implications.

6.2.3.2. Instances of the Direct Argument

Lawrence Shapiro (2000, 2004) imposes a restriction on when a kind is substantively multiply realizable. Namely, two objects are distinct realizers of the same functional kind only if they differ in the properties that are causally relevant to the production of the function that defines that kind (Shapiro 2000, 644). He then goes on to question whether various purported examples of multiply realizable properties satisfy this constraint. For example, he claims that a brain composed of neurons and one composed of silicon chips are different realizations of a mind only in a *trivial* sense, as long as they share the properties (e.g. electrical ones) that are causally relevant to mental functionality. "The difference between standard neurons and silicon neurons, in such a case, is no more interesting relative to their contribution to psychological traits than is the difference between contributions of neurons gray in color and neurons that have been stained purple" (Shapiro 2004, 58).

[138] Clapp's argument in the last sentence for the claim that all properties are multiply realized is clearly wrong. Assume that for any property P, it is not a "brute fact" that P is instantiated. The most this shows is that P is *realized*, not that it is *multiply* realized. In other words, claiming that there are no properties in virtue of which objects instantiate P is claiming that P is not *realized* (not that P is not *multiply* realized), as Clapp claims.

Bechtel and Mundale (1999) marshal similar considerations to cast doubt on the multiple realizability of mental states. They argue that multiple realizability appears plausible only because philosophers mismatch "a broad-grained criterion [of state individuation] (to show sameness of psychological states) with a fine-grained criterion (to differentiate brain states)" (1999, 175). But if the same grain size is used to individuate both mental and brain states, then the multiply realizability of mental states becomes less plausible and one-one mappings can be preserved (ibid., 202). Thus, like Shapiro, they claim that the extent of multiple realization has been overstated because its defenders have appealed to functionally irrelevant chemical and physical properties to individuate brain states, while at the same time individuating psychological states coarsely by abstracting away from such detail.[139]

6.3. An Assumption Shared by the Three Critiques

Each of the three strategies denies a different premise of the multiple realizability argument, but they all can be extended into positive arguments for reducibility that share a common structure and make a key assumption. Namely, they all incorporate the assumption that multiple realizability is a necessary condition for irreducibility. I show this for the three strategies in turn. Then, I argue that some proponents of the multiple realizability argument also make this assumption. I close the section with a discussion of Carl Gillett's take on why the disjunctive and direct strategies fail. Gillett is right to claim that the metaphysics of realization is important, but his specific claims, regarding which features of realization have been overlooked or misunderstood, miss the mark.

---

[139] Thomas Polger (2004) has resurrected what he calls the "Kim-Adams" response to multiple realizaiblity, which claims that mental properties may be realized in physically different systems by physical properties that the systems share. Polger notes that this argument is similar to Shapiro's claim that realizers count as different only if they differ in ways that are relevant to the instantiation of the realized properties (Polger 2004, 11). Further, he thinks that this argument is most successfully deployed in conjunction with the local reduction strategy. For considerations that tell against the "Kim-Adams response," see Block (1980).

The local reduction strategy assumes that its species-specific type-identities will hold if there is a unique realizer of a given mental property in each biological species or structure-type. That is, proponents of this strategy assume that if a property is not multiply realizable, then it is reducible. Thus, this response concedes that multiple realizability is a necessary condition for irreducibility. Since there is allegedly a unique realizer of the given property in each biological species or structure-type, local reductions will be in the offing.[140]

The direct argument puts fairly strong constraints on what counts as genuine multiple realizability and holds that mental properties do not meet these constraints; mental properties are not multiply realizable and thus are reducible. In contrast to the local reduction strategy, Shapiro, Bechtel and Mundale do not argue for species-specific realizers (nor do they deny the existence of a natural property common to all pain experiencers). Rather, they put strong constraints on what counts as genuine multiple realizability and argue that there is no evidence that mental properties meet these constraints. By limiting the scope of multiple realizability, they seek to restrict the number of irreducible properties and clear the way for some form of reductive theory of mind.

The disjunctive strategy focuses on realizer properties; it questions the assumed non-naturalness and heterogeneity of all disjunctive properties and constructs a disjunctive restricted-sense physical property that necessitates the given mental property and is guaranteed to be had by all entities that have the mental property (or at least all nomologically possible ones). The existence of this disjunctive physical

[140] The local reduction strategy is often interpreted as showing that multiple realizability is compatible with reduction or type-identity. However, it is only the allegedly, unnatural, species-independent properties that are multiply realizable according to the eliminative version of this strategy, and even these properties are not multiply realizable according to the conservative version, which incorporates the disjunctive strategy. Note that if the disjunctive strategy is successful, a property may fail to be multiply realizable but still have more than one realizer.

property thus establishes the denial of (2): such mental properties are not multiple realizable. Constructing a disjunctive restricted-sense physical property that is necessarily coextensive with a given mental property is taken to be sufficient to show that the mental property is reducible. Thus, this strategy assumes that unique realizability is sufficient for reducibility—in other words, that multiple realizability is necessary for irreducibility.

The three strategies have the following common structure:[141]

> (Converse of 1) Given *M*, if there is a natural, restricted-sense physical property *P* such that necessarily, for all *x*, *Mx* iff *Px*, then *M* is reducible to *P*. I.e. If *M* is not multiply realizable (uniquely realized), *M* is reducible. I.e. If *M* is irreducible, *M* is multiply realizable.[142]
>
> ¬(2) *M* is not multiply realizable. (There is a *P* such that necessarily, for all *x*, *Px* iff *Mx*.)
>
> So, ¬(5) *M* is reducible.

As I presented the multiple realizability argument above, it uses only the assumption that multiple realizability is sufficient for irreducibility and allows for the possibility of other sources of irreducibility. However, other possible sources of irreducibility are not found in the anti-reductionist literature. The focus is entirely on multiple realization, as reductionist critics have noted (e.g., in the passage from Kim (2003) quoted above). This suggests that proponents of the multiple realizability argument at least at times believed that multiple realizability is also necessary for irreducibility. For example, this assumption is implicit in several passages in Fodor (1974), where he claims that if multiple realizability fails to hold, then the autonomy of special sciences will be merely pragmatic.

---

[141] Of course, the specific mental property that is substituted for '*M*' will vary from strategy to strategy.
[142] Strictly speaking, a property is not multiply realizable if and only if either it is uniquely realized or not realized at all. However, I am assuming that the given property *M* is realized in some way or other.

Further, if subset realization is combined with a theory of properties according to which the causal features of a property are essential to it, then it follows that multiple realizability is necessary for irreducibility. Suppose that a property has only one nomologically possible realizer. According to the theory of properties under consideration, nomological possibility is a special case of metaphysical possibility. All of the causal features that are associated with the unique realizer must also be associated with the realized property. For, any proper subset of these features will not correspond to a property because that subset will not be closed under nomic and metaphysical entailment.[143] Anything that has one of the conditional causal powers from the set associated with the realizer has to have them all. Unique realization gives us property identity: a single set of causal features that define a single property (cf. Shoemaker 2001, 90 and n.26).

Put another way, assuming that the causal features of a property are essential to it, the only way that a realizer can be distinct from a property P that it realizes is if there is a distinct possible realizer for P. For only if there are multiple realizers will be possible for there to be a proper subset of the causal features of each of the realizers that is closed under nomic and metaphysical entailment – and thus defines the realized property. For example, redness is distinct from scarlet because there are causal powers contributed by scarlet and not by red (for example, the power to elicit pecking in birds trained to peck at scarlet objects, but not other shades of red).[144] But the only reason such powers exist is because there are (or at least could be) objects that are red but not scarlet – that is, because red is multiply realized (realizable).

---

[143] One causal power nomically (metaphysically) entails another if and only if it is a consequence of causal laws (a metaphysically necessary truth) that whatever has the first has the second (Shoemaker 2001, 87).

[144] Every causal power contributed by red is also contributed by scarlet (by the subset account of realization). So if it were also the case that every power contributed by scarlet was contributed by red, the sets of causal features associated with red and scarlet would be identical. Hence, according to this account of properties, being red would be identical to being scarlet.

Thus, the assumption that multiple realizability is necessary for irreducibility is shared not only by the critiques on the multiple realization argument but is also plausibly part of the consensus view about what grounds the anti-reductionist consensus.

Recently, Carl Gillett (2003a) has argued against the disjunctive strategy and the direct argument by pointing out that they both depend on the subset account of realization (which he calls the "Flat" or "Standard" account) being correct. He claims that these strategies are unsuccessful against the dimensioned account of realization (which he claims is implicit in the "received" view of the special sciences propounded by Fodor, but not Putnam or Shoemaker). Gillett's general point is that philosophers of mind must pay attention to the metaphysics of realization and that recent critiques of the multiple realizability argument have failed to do so. He claims that "differences over the metaphysical nature of realization are inextricably bound up with broader disputes over the extent of [multiple realizability] and the nature of the special sciences themselves" (2003a, 592).

I think that Gillett's general point is correct, but the metaphysical confusion runs deeper than he imagines. Gillett misidentifies the crux of the dispute and, more importantly, seems to be committed to the mistaken assumption that multiple realizabilty is necessary for irreducibility.

Gillett is right to point out that the Clapp's development of the disjunctive strategy and Shapiro's version of the direct argument assume that the subset account of realization is correct. However, it is not clear that the success of Clapp's and Shapiro's arguments depends on the subset account being true. That is, there are good reasons to doubt that Gillett's dimensioned realization is immune to these strategies. In Chapter 4, I argued that dimensioned realizers still threaten to causally preempt the properties they realize. Even though the dimensioned account does not, as Gillett puts

it, interpret the notion of "causal role playing" literally, realized properties still contribute the powers they do in virtue of the constituents of the dimensioned realizer contributing the powers they do (and not vice versa). As Gillett puts it, "all realizers result in the powers of the realized property, but they may, and often do, contribute no common powers, including those individuative of the realized property" (2003a, 602, italics in original). However "in virtue of" or "resulting in" is spelled out, it is plausible that the disjunctive strategy can develop an extended account of when the causal powers of the realizers appropriately overlap. For example, disjunctive properties need not be individuated by a common set of causal powers; rather, they could be individuated by a set of powers that that all of the disjuncts "result in," that are contributed "in virtue of" their being instantiated.

Similarly, while Shapiro does seem to assume the subset account of realization in his development of the distinction between trivial and substantive realization, this distinction, and the claim that mental properties are only trivially multiply realizable, do not depend on the subset account being correct. The disagreement between Shapiro and his opponents does not ultimately depend on a dispute about the nature of realization but instead seems to turn on empirical questions regarding which molecular and microphysical properties are in fact relevant to the instantiation of mental and other high-level properties.

Gillett does not provide many details about the "received view" of the special sciences, although he claims that it is "an integrated position with commitments about the nature of realization, [multiple realizability], and the special sciences and their laws" (2003, 591 n.2). Apparently, he assumes that the multiple realizability argument will succeed as long as dimensioned realization, but not subset realization, is assumed. He claims that: "The dimensioned account plausibly underpins the received view of the wide extent of [multiple realizability], the heterogeneity of realizers, and,

consequently, its distinctive picture of special sciences and their laws" (2003a, 602-3). Thus, he seems to agree with the consensus view about what grounds the non-reductionist consensus, namely, the multiple realizability of special science properties. In the next section, I argue that the consensus is mistaken. It is not multiple realizability, not even multiple dimensioned realizability, that grounds irreducibility. Rather, it is a broader and metaphysically deeper phenomenon, a many-many relation between mechanisms, which can occur even when a special science property is uniquely realized and of which (substantive) multiple realization is merely one facet.

6.4. Why Multiple Realizability Is Not Necessary for Irreducibility

6.4.1. Multiple Determinativity and One-many Relations Between Sustaining and Integrative Mechanisms

In this section, I use the conceptual resources developed in Chapter 5 in order to argue that multiple realizability is not necessary for irreducibility. Suppose that a determinate mental property *M*, *having an acute pain*, has only one nomologically possible type of total restricted-sense physical structural realizer. Perhaps it is *having Aδ-fibers firing in a nervous system of the appropriate kind*. Suppose that this property is in turn structurally realized by some unique structural realizer involving components of the organism's nervous system. Call this total structural realizer *S* – an immensely complicated structural property of the organism involving properties of and relations between a huge number of molecules and ions. As discussed in the last chapter, *S* provides a sustaining mechanism for *having an acute pain*, a mechanism that can be used to explain how the causal profile of having an acute pain is physically implemented. Also, *S* will be multiply determinative. That is, it will not only realize *M* but will also realize many other kinds of determinate properties of the system in which *M* is instantiated. In the last chapter, I also claimed that the different kinds of properties that are simultaneously realized by *S* correspond to different integrative

mechanisms in which *S* plays a role. Hence, a single sustaining mechanism (corresponding to the components of the structural realizer *S*) can play a role in multiple integrative mechanisms. I claim that this one-many relation between types of sustaining and integrative mechanisms is sufficient for the irreducibility of *M*.
In general, I claim that if a high-level property *Q* has a unique total structural property realizer, *P*, then *Q* is irreducible if (1) *P* is multiply determinative—that is, *P* realizes properties other than *Q* and these other properties are determinates that fall under a different ultimate determinable than *Q*, and (2) the interactions these other properties have with still other properties are captured by different integrative mechanisms of which *P* is a part.

Note that the three reductive strategies discussed above suggest that a one-many relation between properties is not itself sufficient to guarantee that *M* is irreducible. That is, the mere fact that some realizer *P* realizes properties other than *Q* is not enough to block the reduction of *Q* to *P*. For, the reductive strategies are designed to handle one-many relations between properties within a reductionist framework. For example, in cases where a determinate property like *being scarlet* determines a hierarchy of determinables, there is a one-many relation between the determinate and determinables like *being red* and *being colored*. However, ideas drawn from the disjunctive strategy suggest that this one-many relation does not establish that any of the determinables, like *being red*, is irreducible. The determinables and their determinates belong to the same aspect space, and hence will have their relations to other broadly physical properties described by the same type of integrative mechanism and have their instantiations (and the persistence of these) explained by the same type of sustaining mechanism. Hence, they overlap in such a way that one or more of the reductive strategies can be arguably put to use to arrive at a one-one relation between properties. Following the disjunctive strategy, one could

claim that the lower level properties need not be restricted to individual determinates but should include a disjunction of the determinates. Or, following the direct argument, one could claim that the same properties, processes, or mechanisms are relevant to the instantiation of all of the different determinables.

The condition that the properties determined by *P* must fall under a different ultimate determinable than *Q* (i.e. they must be different kinds of property), along with clause (2), is intended to block these strategies. The condition that the realized properties fall under different ultimate determinables and have their interactions described by different integrative mechanisms ensures that there is not the kind of overlap between the simultaneously realized properties as there is between different determinables in a single hierarchy. Since a given realized property will belong to a different aspect space than the other realized properties and the realizer, the reductive claims regarding disjunctions, sameness of causally relevant features, and restriction of reductions to a single structure or species will not apply. The three reductive strategies go wrong by overlooking differences with respect to mechanisms. These differences show that a property can be irreducible even if it is uniquely realized.

In the next subsection, I draw a general moral regarding where debates about reduction went wrong. This will help tie up the last loose end from Chapter 3, by showing how to respond to the explanatory exclusion argument. I also show how claims about mechanisms capture what is distinctive about *substantive* multiple realizability and explain why it has non-reductive implications.[145]

[145] Of course, the extent of multiple determinativity is an empirical question, but there is good reason to think that it will be widespread in the properties studied by psychology and physiology (relative to anatomy and biochemistry) since, roughly speaking, function and form often cross-cut one another.

6.4.2. Reduction, Isomorphism and the Explanatory Autonomy of the Special Sciences

Why is the consensus that multiple realizability is necessary for irreducibility, that unique realizability, a nomologically necessary one-one correspondence, is sufficient for reducibility? I think that answering this question will explain why the debate about reduction has reached an impasse and show how to move beyond it. Briefly, I think it is because the full content of and motivations for reduction have dropped out of the debate; reductionism's implications regarding mechanisms and explanation have largely been ignored. Given this, it seems to be a legitimate reductionist move to "bite the bullet" and settle for extremely local reductions.[146] This move is plausibly only as long as one is ignoring what reductions imply about mechanisms and interactions among properties and focusing solely on nomologically necessary correlations between properties taken in isolation (e.g. at a given instant, cf. n.23).

In an early discussion of possible reasons for thinking that psychophysical supervenience is true, Kim mentions the doctrine of "psycho-physiological isomorphism" which was held, in various forms, by many Gestalt psychologists in the first half of the 20th century (Kim 1982).[147] Although many different claims have been called "psycho-physical isomorphism," I am interested in the broad claim that psychological mechanisms and processes are perfectly mirrored by physical mechanisms and processes. This claim is not only a feature of early 20th century psychological theories. It is present in type-identity theories, like that developed by U.T. Place and J.J.C. Smart, and it is explicitly assumed by some work in

[146] Cf. Kim: "In the worst-case scenario in which there is wildly heterogeneous multiple realization everywhere among the humans, and for the same individual over time, there still would be *structure*-specific biconditional laws (if psychology is indeed physically realized), and there still would be perfectly good local reductions, even if they are only for single individuals at a particular moment of their lives" (1998, 94).

[147] Although gestalt psychologists were not identity theorists, many of them were reductionists in that they thought that psychological changes were perfectly mirrored by changes in physiological fields.

contemporary cognitive science (see the discussion of "analytic isomorphism" and the "bridge locus" in (Pessoa et al. 1998)). Further, it seems to be implicit in the method of subtraction employed in fMRI and PET brain imaging, which often assumes that the structure of individual cognitive components of a complex cognitive task will be isomorphic to the structure of neural activation (see, e.g., Friston et al. 1996, Sidtis et al. 1999).

If all special sciences reduce to physics, then a broader claim must be true: all psychological and other high-level causal interactions must be perfectly mirrored by physical causal interactions – what might be called "omni-physical isomorphism." This is an integral part of reductionism, for it is required if low-level physical causal interactions are to give a complete account of world (i.e. if all natural properties are identical to natural physical properties).

Debates about reduction have concerned a weaker claim that affirms the existence of lawlike correlations between physical and mental properties (see e.g., Kim 1982, 178). Of course, these local lawlike correlations are not sufficient for reduction, but the idea is that they can easily be "enhanced" into identities, as Kim later put it (1998, 97). However, this will be possible only if omni-physical isomorphism is true, since every identity relation is an isomorphism. These debates have focused on only one way in which this isomorphism could fail: if mental properties do not map uniquely onto the domain of physical properties, i.e. the problem that substantive multiple realizability poses for reducibility – if it is not possible to find a single realizer to match up with a given realized property. The three reductive strategies discussed above attempt to defuse this problem. Whether they are successful is debatable. But even if they are, the omni-physical isomorphism thesis will also be false if physical properties do not map uniquely onto the domain of high-level properties, i.e. the problem that multiple determinativity

poses for reducibility. These strategies do not rule out a single realizer property realizing many kinds of high-level property (i.e. multiple determinativity). What has been overlooked by all parties in the debate is the fact that the various reductionist strategies that attempt to get around the first problem all flounder when it comes to the second problem. Other implications of reductionism – specifically, that the structure of mechanisms involving mental properties will be perfectly mirrored by the structure of mechanisms involving restricted-sense physical properties – have dropped out of the debate. Basing irreducibility solely on multiple realizability would be compatible with this reductionist implication. For, multiple realizability is logically compatible with unique determinativity: if each of the various realizers of a high-level property *Q* realizes only *Q*, then each realizer will map uniquely onto the domain of high-level properties.[148]

The fact that a mental property *M* has one possible restricted-sense physical structural realizer *S* does not entail that the psychological mechanisms of which *M* is a part will be perfectly mirrored by the mechanisms in which the restricted-sense physical structural realizer *S* figures. As discussed above, *S* will be multiply determinative, and it will be a part of multiple integrative mechanisms that capture the relations between the other properties *S* realizes and still other high-level properties. If we replace *M* with *S*, then we conflate these integrative mechanisms. The explanation in terms of the realizer *S* may contain *more* information since it realizes many properties and plays a role in a corresponding number of sets of causal

[148] Among other places, this assumption is reflected in Putnam's (1982) claim that the realism he formerly espoused is committed to there being only one set of natural kinds – namely, fundamental physical ones – and that these capture all of the world's causal structure. If this were so, then each occurrence of a realized property could be replaced by its unique realizer. The view defended here challenges this reductionist assumption. It holds that there are multiple kinds (or systems) of natural kinds and that natural special science kinds are not natural restricted-sense physical kinds, and vice versa.

interactions, but not all of this information is relevant to the mental or behavioral effect at issue.

Physical properties are unable to sort out or distinguish between the many different sets of high-level causal interactions in which *S* participates; information regarding these causal interactions thus gets garbled or confused at the microphysical level. This supports the autonomy of special science explanations that cite these properties. Contra Kim (1989, 241, 249), special science explanations may be deeper, more detailed, and theoretically more fecund because they isolate integrative mechanisms in which *S* enters that are relevant to *M* from those that are irrelevant. "Looking down" to the sustaining mechanism (as opposed to "looking up" at the integrative mechanisms, as was done in the previous paragraph), the fact that *S* is multiply determinative implies that there is no sustaining mechanism that explains the instantiation and persistence of *only M*. The same restricted-sense physical type of sustaining mechanism will also explain the persistence of instantiations of the other properties realized by *S*. For this reason, *S* does not capture the causal or modal essence of *M*. There are more distinctions in the world than are captured by restricted-sense physical properties.

This facet of multiple determinativity is relevant to another charge made by Kim regarding the supposed deficiency of special science explanations (1989, 251). Namely, the broadly physical account may be more theoretically systematic than restricted-sense physical explanations, since, in cases of multiple determinativity, there is no well-defined restricted-sense physical sustaining mechanism that is unique to a given broadly physical property. Further, there is no way to separate out the parts of the structural realizer that are directly responsible for the instantiation and persistence of *having an acute pain* from those that are directly responsible for the

instantiation and persistence of the other properties that *S* realizes (say, *having functioning neurons* or *having a certain elasticity*).[149]

We are now also in a position to see why substantive multiple realizability is sufficient for irreducibility. Suppose that *M*, *having an acute pain*, is substantively multiply realizable. That is, suppose that there are many structural realizers, $P_1, P_2, \ldots$ of *M*, each of which corresponds to a different type of restricted-sense physical sustaining mechanism that explains the instantiation and persistence of *M*. But each of these sustaining mechanisms and structural realizers plays the same role in the various integrative mechanisms in which *M* figures. Each of the P_is will be related to other restricted-sense physical properties (states of affairs types) that realize the other properties with which *M* enters into causal relations. Again, irreducibility does not arise simply because there is a one-many relation that holds between mental and physical properties. Rather, in this case, irreducibility is grounded in a one-many relation between integrative and sustaining mechanisms

Attempting to reduce *M* to any of the P_is would lead one to mistakenly conclude that the details of the sustaining mechanism corresponding to P_i were essential to sustaining *M* and accounting for its relations to other mental (and other broadly physical) properties. The existence of the other realizers shows that this conclusion is false; other types of mechanism can sustain *M* and account for its relations to other properties. Further, the framework explains *why* the disjunctive strategy is unsuccessful against substantive multiple realizability. The disjunction of the P_is is not a natural restricted-sense physical property because it is involved in distinct kinds of restricted-sense mechanisms. Many defenders of the multiple

[149] Of course, if these other properties are multiply realizable (as is likely), this will raise further problems for the reductionist position. For, the relation between properties will now be many-many. But we can assume for the sake of argument (and probably contrary to fact) that there is only one way to realize these other properties.

realizability argument have claimed that such a disjunction exhibits causal unity only at the mental level (e.g. Antony and Levine 1997, 90). The framework presented here explains this claim. Note, however, that the local-reduction strategy still poses a problem for the multiple realizability argument, unless there are different kinds of mechanisms that sustain *M* within the same species.

Thus, the mechanistic realization framework encompasses traditional claims of multiple realizability, but it also captures the equally important claims about mechanisms that are associated with multiple determinativity, which the consensus view has overlooked. Taken together these claims show that the relation between high-level and restricted-sense physical mechanisms will in general be many-many; they will cross-cut one another.

6.5. Three objections

In this section I respond to three objections, which will help to clarify what I take to be the basis of irreducibility.

First, it might seem that my account leads to the implausible conclusion that *all* properties are irreducible. For example, it may seem that any structural realizer of the property *having mass of 10kg* will also be multiply determinative and hence *having mass of 10kg* will be irreducible. But *having mass of 10kg* is plausibly *not* irreducible in the way that, say, mental properties, like *having an acute pain*, are. I think that this difference between properties like *having mass of 10kg* and mental properties like *having an acute pain* is explained by a difference in their sustaining mechanisms. *Having mass of 10kg* may be instantiated in any old lump of matter. Hence, its structural realizer type (assuming there is only one) is not complex enough to determine other kinds of property. The sustaining mechanism for any determinate mass will not involve complex interactions between different kinds of restricted-sense physical properties. Rather, it will merely involve the aggregation or sum of the

requisite number of microphysical entities. The same kind of mechanism is required at any scale. Following Millikan's suggestion, we can say that looking more closely at the structural realizer of (different instantiations of) *having mass of 10kg* only reveals "what stuff it is made of" not how the sustaining mechanism works. However, looking more closely at the structural realizer of (different instantiations of) *having an acute pain* can reveal how the mechanism works.

Although not all properties will be irreducible, it is likely that some chemical and macrophysical properties, perhaps *being an electrical conductor* and *being rigid*, will be, if their structural realizers are multiply determinative (or if they are substantively multiply realizable) (see Block (1997) and (2003) for brief, but conflicting, discussions of the irreducibility of rigidity). Some philosophers may see this as an unwelcome consequence. I think that any uneasiness about such irreducible physical properties derives from an inadequate view of reduction which claims that properties are irreducible only if they figure only in the laws or generalizations of a special science and do not appear in any low-level, physical laws. If properties like *being rigid* turn out to be irreducible, this merely emphasizes that the kind of irreducibility I am interested in is not grounded in relations between scientific disciplines or in which generalizations humans happen to find illuminating or interesting.[150]

Turning to the second objection, I have claimed that the causal mechanisms associated with high-level realized properties and their restricted-sense physical realizers cut across one another. But Kim has argued that claims about cross-classification violate physicalism. He claims that:

> To say a given taxonomic system cross-classifies another must mean something like this: there are items that are classified in the same way,

[150] However, as I discuss below, mismatch between mechanisms can explicate some of the pragmatic and epistemic claims that non-reductionists have made.

> and cannot be distinguished, by the second taxonomy (that is, indiscernible in respect of properties recognized in this taxonomy) but that are classified differently according to the first taxonomy (that is, discernable in respect of properties recognized in that taxonomy), and perhaps vice versa. That is, a taxonomy cross-classifies another just in case the former makes distinctions that cannot be made by the latter (and perhaps also conversely). But then this means that the first taxonomy fails to supervene on the second, and [cross-classification] when understood this way must come to the denial of supervenience of higher-order properties. If mental properties and biological properties cross-classify basic physical properties, then cannot supervene on the latter. … I think that in talking about cross-classifying, [Horgan] may simply be referring to the familiar claim of multiple realizability of higher-order properties in relation to basic physical properties. [But,] the idea that mental properties are physically realized, whether multiply or uniquely, logically entails the supervenience thesis. If you accept physical realizationism, as I believe Horgan does, you cannot at the same time hold the "cross-classification" thesis, at least in the present sense" (Kim 1998, 68-9).

I believe the cross-classification claim is better understood as a claim about multiple determinativity, not multiple realizability. The high-level properties simultaneously realized by a multiply determinative physical realizer make distinctions that cannot be made by the single physical realizer. They fall into separate integrative mechanisms which are not distinguishable in restricted-sense physical terms. But because multiple determinativity is not a modal claim in the way that multiple realizability is, this is not incompatible with physicalism. Cross-cutting or cross-classification is not a modal claim about differences between worlds that would violate the supervenience of the high-level on the restricted-sense physical. Rather, it is a claim about the patterns in this world that are captured by different mechanisms. A single restricted-sense physical sustaining mechanism, corresponding to the components of a structural realizer, cannot differentiate between all of the high-level integrative mechanisms involving the properties it realizes. It collapses them all into a single mechanism, thereby conflating causal structure that is present in the world.

Finally, one might simply to deny that natural microphysical structural realizers like *S* are multiply determinative. That is, one might allege that my argument is mistaken because it focuses on a natural structural realizer type that is too broad. By singling out certain aspects or parts of the structural realizer, this response claims, we can obtain a "stripped down" realizer that will not be multiply determinative, a natural structural realizer that is specific to a given maximally determinate high-level property, realizing only it (and derivatively all of the properties that it subset realizes). Put another way, such a response might claim that once one investigates the scientific details of the sustaining mechanism (the components of the total structural realizer *S*) one will find parts of that mechanism that contribute directly to the instantiation of *having an acute pain* but that do not contribute directly to the instantiation of any other realized property.

Consider the second version of this response. Far from allowing one to identify a natural realizer that is specific to, say, *having an acute pain*, I think that investigating the scientific details will only show that the natural total realizer is even more complicated and realizes an even greater variety of properties than it initially appeared to. In this chapter, I have been indulging the philosopher's conceit that the physiological realizer of *having an acute pain* is something as simple as *having Aδ-fibers firing in a nervous system of the appropriate kind*. Even setting aside the fact that nothing has been said about which nervous systems are of the "appropriate kind," the core realizer of *having an acute pain* cannot be anything as simple as *having Aδ-fibers firing*. The nociceptor system in humans is highly plastic, and under certain conditions "other receptor types which are normally associated with touch acquire the capacity to evoke pain" (Meyer et al. 1994).

This suggests that pain is multiply realizable in the standard sense, which implies that any core realizer which is common to all humans is going to be a very

complicated disjunctive property. But it also suggests that a single instance of *having an acute pain* may be realized by different fibers in the nervous system at different times.[151] Further, under certain conditions, parts of the nervous system may be recruited to underwrite the instantiation of an instance of *having an acute pain* that would not be sufficient (even embedded in the right kind of system) to realize pain by themselves. Such cases are not standard examples of multiple realization. There are not multiple distinct realizers of pain in such cases. Rather, the realizer "leaks out" from the standard nociceptor system into other parts of the nervous system. (This, incidentally, emphasizes the importance of processes/mechanisms in an adequate realization-based account of physicalism. The sustaining mechanism for a single token of pain may change over time; the stability and persistence of an instance of pain may be underwritten by great fluidity at the neural level.) In a similar vein, there is apparently no such thing as a "pain system" in the human nervous system; the "parts" of the realizer that are directly relevant to the instantiation of *having an acute pain* will include neurological structures that are also directly relevant to the instantiation of states like attention and mood (Wall and Jones 1992).

This leads to a reply in terms of the first version of the response. There are not parts of a structural realizer of *having an acute pain* that are directly relevant to the instantiation of *having an acute pain* that are not directly relevant to the instantiation of other properties that are realized by that realizer. All parts of the natural realizer that are directly relevant to the implementation of one realized property are directly relevant to the instantiation of at least one other realized property. Given the assumption of unique realization, these realized properties are always found together.

[151] Cf. Antony and Levine's argument that "the arbitrariness inherent in the disjunctive strategy isn't just in the collecting of disjunct properties; *it is in the construction of the disjunct-properties themselves*" (1997, 90, italics in original). Helen Steward made a similar point in her comments on a version of some of this material that I read at the Oxford Philosophy Graduate Conference in November 2005.

The objection I am considering claims that we can find a structural realizer that uniquely realizes a given high-level property – so that there will be a one-one correspondence between that structural realizer and the high-level property. I claim that once we identify something that approximates a total structural realizer for, say, *having an acute pain* – one that is *sufficient* for the instantiation of acute pain – any part of the realizer that is directly relevant to acute pain's instantiation will also be directly relevant to the instantiation of other high-level properties.

My reply can be put in the form of a dilemma. Either the realizer is "stripped down" so much, in an attempt to get a restricted-sense physical type that is specific to the mental property, that it is either no longer is a total realizer – is no longer sufficient for the instantiation of the realized property – or is no longer a natural restricted-sense physical type. Or, as more of the nervous system is added back in to ensure that the realizer is total, the realizer ceases to be specific to the mental property in question (realizing only it and no other high-level properties).

Consider another example. Diamonds instantiate the property *having a high melting point*. The covalent sigma bonds in diamond are directly relevant to its having a high melting point. (They are plausibly the only parts of the molecular structure that are directly relevant.) However, these bonds are *also* directly relevant to diamond's instantiating the property *having poor electrical conductivity* (because all of the electrons are tightly bound in the tetrahedral structure of the sp3 hybridized bonds). I claim that this is not simply an artifact of nervous systems and diamonds. Rather, it is common to all articulated macro-properties – those which require a particular complicated system (e.g. an organ system or organism) for their realization (and not merely a big glob of matter). This stems from the fact that the components of the sustaining mechanism – the properties of and relations that make up the structural realizer – belong to a different aspect space than the realized property. These

components determine the realized property only *in concert* with one another. The interactions between these components produce a *suite* of novel aspects and powers (corresponding to the many realized properties) that are not produced by those components in isolation. So we cannot just single out the restricted-sense physical aspects that are directly relevant to the instantiation of one of these articulated macro-properties as we could with molar macro-properties like mass and density.

6.6. Explicating Claims Made by Other Non-reductive Physicalists

Claims about multiple determinativity and mechanisms also explain two initially puzzling footnotes from papers that are cited as *loci classici* of the multiple realizability argument. Discussing these passages will provide further clarification of the metaphysical basis of the autonomy of high-level explanations and show why Yablo's claim that a cause must be proportional to or commensurate with its effect is not merely a pragmatic constraint.

As discussed above, the standard interpretation of Putnam's "square peg/round hole" argument is that it merely an application of the multiple realizability argument. This interpretation is supported by Putnam's presentation of the argument in the main body of the paper. However, I believe this interpretation is mistaken; it misidentifies the reason that lower level accounts (those in terms of narrowly physical properties) are not explanatory. It need not be multiple realizability, but may also be multiple determinativity, that grounds these explanatory facts.

After presenting the square peg and round hole example, Putnam makes the following remark in a footnote:

> Even if it were not physically possible to realize human psychology in a creature made of anything but the usual protoplasm, DNA, etc., it would still not be correct to say that psychological states are identical with their physical realizations. For, as will be argued below, such an identification has no *explanatory* value *in psychology*. (1975b, 293)

Elliott Sober notes that this remark implies that Putnam must think that "the virtue of higher-level explanations does not reside in their greater generality" (1999, 549 n. 8). For, if *P* has one possible physical realizer, *R*, then *P* and *R* belong to exactly the same objects. Sober speculates that Putnam would say that citing *R* in an explanation provides "extraneous information," whereas citing *P* does not.

Similar ideas appear in Philip Kitcher's (1984). Kitcher claims that a derivation of the general principles of classical genetics from molecular genetics is not explanatory because "in charting the details of the molecular rearrangements the derivation would only blur the outline of a simple cytological story, adding a welter of irrelevant detail" (1984, 347). The natural reductionist response is to claim that this assumes an overly subjective view of explanation. Beings with limited cognitive capacities like *us* may get lost in the detail, but there is nothing *objectively* deficient about such "messy" molecular derivations.

In response to this line of thought, Kitcher claims that molecular derivations objectively fail to explain because the natural kind *PS-process* ("PS" stands for "pair separation"), to which meiosis belongs, cannot be identified as a natural kind from the molecular point of view. In a footnote in which he refers to Putnam's discussion of the cubical peg/round hole example, Kitcher elaborates on this point:

> It would be tempting to think that the independence of the "higher level structural features" in Putnam's example and in my own can be easily established: one need only note that there are worlds in which the same feature is present without any molecular realization. So, in the case discussed in the text, PS-processes might go on in worlds where all objects were perfect continua. But although this shows that PS-processes form a kind which could be realized without molecular reshufflings, we know that all *actual* PS-processes do involve such reshufflings. The reductionist can plausibly argue that *if* the set of PS-processes with molecular realizations is itself a natural kind, then the explanatory power of the cytological account can be preserved by identifying meiosis as a process of this narrower kind. Thus the crucial issue is not whether PS-processes form a kind with nonmolecular

> realizations, but whether those PS-processes which have molecular realizations form a kind that can be characterized from the molecular point of view. (1984, 350 n. 20)

These two passages suggest that multiple realizability is not necessary for blocking reductive explanation, in the sense of the reducing theory preserving all of the explanatory power of the reduced theory. Instead, the key issue seems to be something about whether explanations are available at lower "levels of organization" or "points of view." At least for Kitcher, this is tied up with the question of whether kinds that are natural at one level can be identified as natural kinds at a lower level. If Kitcher is right about what the "crucial issue" is, then it is possible that a natural mental kind is not a natural molecular kind, *even if it is realized by a single molecular structure*.[152] I have argued that this possibility is actualized and that claims about cross-cutting mechanisms explain why some high-level kinds cannot be completely characterized from the restricted-sense physical "point of view."

The realization framework presented here thus clarifies complex and contentious issues about explanation and natural kinds that, taken by themselves, are too easily dismissed as merely pragmatic, epistemic, or overly anthropocentric. The fact that structural realizers are multiply determinative provides a metaphysical basis for their claim that reductive explanations miss generalizations by including irrelevant detail and obscuring causally relevant features. That is, multiple determinativity provides a basis for the claims that (i) there is detail in the structural realizer that is irrelevant to a given high-level effect, and (ii) explanatory features relevant to the

[152] However, even Kitcher himself does not seem to recognize this possibility in the main text of his article. There he argues that PS-processes do not form a molecular kind by appealing to familiar considerations of multiple realizability. He writes: "PS-processes are heterogeneous from the molecular point of view. There are no constraints on the molecular structures of the entities that are paired or on the ways in which the fundamental forces combine to pair them and to separate them" (1984, 349). "PS-processes are realized in a motley of molecular ways" (ibid., 350). Thus, he seems to think that the crucial issue is not whether PS-processes are in general multiply realizable (perhaps non-molecularly). Rather, it is whether those PS-processes that are molecularly realized are themselves "heterogeneous" or "motley" at the molecular level. This misses the important point he seemed to be moving toward in the footnote quoted above.

explanation of that effect are obscured or conflated at the microphysical level and are only "visible" at the higher level.

Return to the acute pain example from above. Contra Kitcher, the reductive explanation need not fail because *M* is *realized in* "a motley of neural or electronic ways" (see note 152). Rather, it may also fail because a motley assortment of properties is *realized by* the single neural structure that uniquely realizes *M*. Citing the structural realizer in an explanation is equivalent to citing the disjunction of all the properties that are realized by it: the acute pain along with, e.g., the mass, color, and myelin content of the system constituted by the structural realizer, its conductivity, etc.[153] If we want to explain an effect that is characteristic of pain (wincing, moaning, etc.), an explanation that cites the structural realizer will be objectively worse because it includes all of these irrelevant properties and their corresponding integrative mechanisms. By contrast, by citing the realized property we single out only those features that are causally relevant to the effect in question.

As noted above, it is important that different *kinds* of properties be realized. If this were not the case, it is not clear that any *irrelevant* detail would be included. For example, if the many realized properties were a hierarchy of determinables, each would be less specific than the realizer *in the same way* (Gasper (1992) provides this kind of example.) They would be distinguished by different ranges of the *same* causally relevant aspects, not by different *kinds* of causally relevant aspects (such as those that characterize *having an acute pain*, *having functioning neurons*, and *having a certain conductivity*).

Finally, Yablo's commensurability intuition—roughly, that causes "should incorporate a good deal of causally important material but not too much that is causally unimportant" (1992, 274)—is obviously similar to Putnam's claim that shape

[153] On this point, see Gasper (1992).

and rigidity, and not the detailed molecular realizer of these properties, explain why the square peg did not go through the round hole. Yablo develops four conditions—collectively known as proportionality—which he claims capture the commensurability intuitions. However, like Putnam and Kitcher (except in the two footnotes cited above), he defends the claims that mental properties are distinct from restricted-sense physical properties and are sometimes causes by appealing to their multiple realizability (1992, 278). Further, he defends the proportionality constraints and their application in particular cases by appealing to modal intuitions and to "what most people would say" (ibid.). This makes it look as if proportionality is merely a pragmatic constraint on the causal explanations that we find more illuminating or interesting.

If the discussion in this chapter is correct, then the causal efficacy of mental properties and the autonomy of explanations that utilize them are not ultimately grounded in their multiple realizability, but rather in the fact that the relation between high-level and realizing mechanisms is, in general, many-many. Substantive multiple realizability is simply one facet of this phenomenon, corresponding to the one-many between a single high-level integrative mechanism and realizing sustaining mechanisms. However, the other facet, the converse *many-one* relation between high-level integrative mechanisms and a single sustaining mechanism, fares equally well at establishing that high-level properties can be cited as causes. Further, this realization framework shows why proportionality constraints are not merely a result of our explanatory or conceptual schemes but are grounded in objective features of the world.

REFERENCES

Antony, Louise. (1999a) "Making Room for the Mental: Comments on Kim's `Making Sense of Emergence'." Philosophical Studies. 95: 37-44.

---. (1999b) "Multiple Realizability, Projectiblity, and the Reality of Mental Properties." Philosophical Topics. 26: 1-24.

Antony, Louise M. and Joseph Levine. (1997) "Reduction with Autonomy." Philosophical Perspectives. 11: 83-105.

Bailey, Andrew. (1999) "Supervenience and Physicalism." Synthese. 117: 53-73.

Baker, Lynne Rudder. (1993) "Metaphysics and Mental Causation." in Mental Causation. Ed. John Heil and Alfred Mele. Oxford: Clarendon. 75-95.

Baudis, Laura. (2006) "Dark Matter Searches." International Journal of Modern Physics A. 21(8-9): 1925-1937.

Bealer, George. (1997) "Self-Consciousness." Philosophical Review. 106: 69-117.

Bechtel, William and Jennifer Mundale. (1999) "Multiple Realizability Revisited: Linking Cognitive and Neural States." Philosophy of Science. 66(2): 175-207.

Beckermann, Ansgar. (1992) "Introduction - Reductive and Nonreductive Physicalism." in Emergence or Reduction? Ed. Hans Flohr Ansgar Beckermann, and Jaegwon Kim. Berlin: W. de Gruter. 1-21.

Bennett, Karen. (2003) "Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It." Noûs. 37(3): 471-497.

---. (forthcoming) "Exclusion Again." in Being Reduced. Ed. Jakob Hohwy and Jesper Kallestrup. Oxford: Oxford UP.

Bickle, John. (1998) Psychoneural Reduction: The New Wave. Cambridge, MA: MIT Press.

Block, N.R. and J.A. Fodor. (1972) "What Psychological States Are Not." Philosophical Review. 81(2): 159-181.

Block, Ned. (1980) "Introduction: What is Functionalism?" in Readings in Philosophy of Psychology. Ed. Ned Block. Cambridge, MA: Harvard UP. 170-184.

---. (1997) "Anti-Reductionism Slaps Back." Philosophical Perspectives. 11: 107-132.

---. (2003) "Do Causal Powers Drain Away?" Philosophy and Phenomenologial Research. 67(1): 133-150.

Bontly, Thomas D. (2002) "The Supervenience Argument Generalizes." Philosophical Studies. 109: 75-96.

Boyd, Richard N. (1980) "Materialism Without Reductionism: What Physicalism Does Not Entail." in Readings in Philosophy of Psychology. Ed. Ned Block. Cambridge, MA: Harvard UP. 67-106.

---. (1999) "Kinds, Complexity and Multiple Realization: Comments on Millikan's 'Historical Kinds and the Special Sciences'." Philosophical Studies. 95: 67-98.

Burge, Tyler. (1993) "Mind-Body Causation and Explanatory Practice." in Mental Causation. Ed. John Heil and Alfred Mele. Oxford: Clarendon. 97-120.

Cartwright, Nancy. (1999) The Dappled World. Cambridge: Cambridge UP.

---. (2004) "Causation: One Word, Many Things." Philosophy of Science. 71: 805-819.

Chalmers, David. (1994) "On Implementing a Computation." Minds and Machines. 4(4): 391-402.

---. (1996) The Conscious Mind: In Search of a Fundamental Theory. New York: Oxford UP.

Clapp, Lenny. (2001) "Disjunctive Properties: Multiple Realizations." Journal of Philosophy. 98(3): 111-136.

Crane, Tim. (1995) "The Mental Causation Debate." Proceedings of the Aristotelian Society. Supp. 69: 211-236.

Crane, Tim and D.H. Mellor. (1990) "There Is No Question of Physicalism." Mind. 99(394): 185-206.

Craver, Carl F. (2001) "Role Functions, Mechanisms, and Hierarchy." Philosophy of Science. 68: 53-74.

Crook, Seth and Carl Gillett. (2001) "Why Physics Alone Cannot Define the 'Physical': Materialism, Metaphysics, and the Formulation of Physicalism." Canadian Journal of Philosophy. 31(3): 333-360.

David, Marian. (1997) "Kim's Functionalism." Philosophical Perspectives. 11: 133-148.

Davidson, Donald. (1980) "Mental Events." in Essays on Actions and Events. Oxford: Clarendon. 207-224.

---. (1993) "Thinking Causes." in Mental Causation. Ed. John and Alfred Mele Heil. Oxford: Clarendon. 3-17.

Dennett, Daniel. (1991) "Real Patterns." Journal of Philosophy. 88(1): 27-51.

Egan, Frances. (1991) "Must Psychology Be Individualistic?" Philosophical Review. 100: 179-203.

---. (1995) "Computation and Content." Philosophical Review. 104(2): 181-203.

Ehring, Douglas. (1996) "Mental Causation, Determinables and Property Instances." Noûs. 30(4): 461-480.

Ellis, Brian. (2005) "Universals, The Essential Problem and Categorical Properties." Ratio. 18(4): 462-472.

Endicott, Ronald. (1994) "Constructival Plasticity." Philosophical Studies. 74: 51-75.

---. (1998) "Many-Many Mappings and World Structure." Americal Philosophical Quarterly. 35: 261-280.

Feigl, Herbert. (1958) "The 'Mental' and the 'Physical'." in Minnesota Studies in the Philosophy of Science. Ed. Herbert Feigl, Michael Scriven and Grover

Maxwell. Minneapolis. Vol. 2, Concepts, Theories, and the Mind-Body Problem. 370-497.

Fodor, Jerry. (1974) "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)." Synthese. 28: 97-115.

---. (1987) Psychosemantics. Cambridge, MA: MIT Press.

Francescotti, Robert. (2002) "Understanding Physical Realization (and what it does not entail)." Journal of Mind and Behavior. 23(3): 279-292.

Friston, K.J., et al. (1996) "The Trouble with Cognitive Subtraction." Neuroimage. 4: 97-104.

Gasper, Philip. (1992) "Reduction and Instrumentalism in Genetics." Philosophy of Science. 59: 655-670.

Gillett, Carl. (2002) "The Dimensions of Realization: A Critique of the Standard View." Analysis. 62(4): 316-323.

---. (2003a) "The Metaphysics of Realization, Multiple Realizability, and the Special Sciences." Journal of Philosophy. 100(11): 591-603.

---. (2003b) "Non-Reductive Realization and Non-Reductive Identity: What Physicalism Does Not Entail." in Physicalism and Mental Causation. Ed. Sven Walter and Heinz-Dieter Heckmann. Exeter: Imprint Academic. 31-57.

Gillett, Carl and Bradley Rives. (2001) "Does the Argument from Realization Generalize? Responses to Kim." Southern Journal of Philosophy. 39: 79-98.

Glennan, Stuart. (2002) "Rethinking Mechanistic Explanation." Philosophy of Science. 69: S342-S353.

Goldman, Alvin I. (1969) "The Compatibility of Mechanism and Purpose." Philosophical Review. 78(4): 468-482.

Hare, R.M. (1952) The Language of Morals. Oxford: Clarendon.

Hawthorne, J.P. (2002) "Blocking Definitions of Materialism." Philosophical Studies. 110: 103-113.

Heil, John. (1992) The Nature of True Minds. Cambridge: Cambridge UP.

---. (1999) "Multiple Realizability." American Philosophical Quarterly. 36: 189-208.

---. (2003) "Multiply Realized Properties." in Physicalism and Mental Causation. Ed. Sven Walter and Heinz-Dieter Heckmann. Exeter: Imprint Academic. 11-30.

Hellman, Geoffrey. (1985) "Determination and Logical Truth." Journal of Philosophy. 82: 607-616.

Hellman, Geoffrey P. and Frank Wilson Thompson. (1975) "Physicalism: Ontology, Determination, and Reduction." Journal of Philosophy. 72: 551-564.

Hempel, Carl G. (1980) "The Logical Analysis of Psychology." in Readings in Philosophy of Psychology. Ed. Ned Block. Cambridge, MA: Harvard UP. 14-23.

Hill, Christopher. (1991) Sensations. Cambridge: Cambridge UP.

Hirsch, Eli. (1980) The Concept of Identity. New York: Oxford UP.

Hodges, Andrew. (1983) Alan Turing: The Enigma. New York: Simon and Schuster.

Horgan, Terence. (1989) "Mental Quausation." Philosophical Perspectives. 3: 47-76.

---. (1993) "From Supervenience to Superdupervenience: Meeting the Demands of a Material World." Mind. 102(408): 555-586.

---. (1997) "Kim on Mental Causation and Causal Exclusion." Philosophical Perspectives. 11: 165-184.

Horwich, Paul. (1998) Meaning. Oxford: Oxford UP.

Jackson, Frank. (1998) From Metaphysics to Ethics. Oxford: Clarendon.

Johnson, W.E. (1921) Logic: Part I. Cambridge: Cambridge UP.

Kim, Jaegwon. (1966) "On the Psycho-Physical Identity Theory." American Philosophical Quarterly. 3: 227-235.

---. (1976) "Events as Property Exemplifications." in Action Theory. Ed. Myles Brand and Douglas Walton. Dordrecht. 159-177. (Reprinted in Kim (1993c) pp. 33-52.)

---. (1980) "Physicalism and the Multiple Realizability of Mental States." in Readings in Philosophy of Psychology. Ed. Ned Block. Cambridge, MA: Harvard UP. 234-236. (From "Phenomenal Properties, Psychophysical Laws, and the Identity Theory." Monist (1972) 56: 177-192.)

---. (1982) "Psychophysical Supervenience." Philosophical Studies. 41: 51-70. (Reprinted in Kim (1993c) pp. 175-193.)

---. (1984) "Epiphenomenal and Supervenient Causation." Midwest Studies in Philosophy. 9: 257-270. (Reprinted in Kim (1993c) pp. 92-108.)

---. (1988) "Supervenience for Multiple Domains." Philosophical Topics. 16: 129-150. (Reprinted in Kim (1993c) pp. 109-130.)

---. (1989) "Mechanism, Purpose and Explanatory Exclusion." Philosophical Perspectives. 3: 77-108. (Reprinted in Kim (1993c) pp. 237-264.)

---. (1992) "Multiple Realization and the Metaphysics of Reduction." Philosophy and Phenomenologial Research. 52: 1-26. (Reprinted in Kim (1993c) pp. 309-335.)

---. (1993a) "The Nonreductivist's Troubles with Mental Causation." in Mental Causation. Ed. John Heil and Alfred Mele. Oxford: Clarendon. 189-210. (Reprinted in Kim (1993c) pp. 336-357.)

---. (1993b) "Postscripts on Mental Causation." in Supervenience and Mind. Cambridge: Cambridge UP. 358-367.

---. (1993c) Supervenience and Mind. Cambridge: Cambridge UP.

---. (1998) Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation. Cambridge, MA.

---. (1999) "Making Sense of Emergence." Philosophical Studies. 95: 3-36.

---. (2003) "Blocking Causal Drainage and Other Maintenance Chores with Mental Causation." Philosophy and Phenomenologial Research. 67(1): 151-175.

---. (2005) Physicalism, Or Something Near Enough. Princeton: Princeton UP.

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

Kitcher, Philip. (1984) "1953 and All That. A Tale of Two Sciences." Philosophical Review. 93(3): 335-373.

Latham, Noa. (2003) "What Is Token Physicalism?" Pacific Philosophical Quarterly. 84(3): 270-290.

LePore, Ernest and Barry Loewer. (1989) "More on Making Mind Matter." Philosophical Topics. 17: 175-191.

Lewis, David. (1966) "An Argument for the Identity Theory." Journal of Philosophy. 63(1): 17-25.

---. (1970) "How to Define Theoretical Terms." Journal of Philosophy. 67(13): 427-446.

---. (1972) "Psychophysical and Theoretical Identifications." Australasian Journal of Philosophy. 50: 249-258. (Reprinted in Lewis (1999) pp. 248-261.)

---. (1980) "Review of Putnam." in Readings in Philosophy of Psychology. Ed. Ned Block. Cambridge, MA: Harvard UP. 232-233. (From "Review of Art, Mind, and Religion" in Journal of Philosophy. (1969) 66: 23-35.)

---. (1983) "New Work for a Theory of Universals." Australasian Journal of Philosophy. 61: 343-377. (Reprinted in Lewis (1999) pp. 8-55.)

---. (1986) On the Plurality of Worlds. Oxford: Blackwell.

---. (1994) "Reduction of Mind." in A Companion to Philosophy of Mind. Ed. Samuel Guttenplan. Oxford. (Reprinted in Lewis (1999) pp. 291-324.)

---. (1999) Papers in Metaphysics and Epistemology. Cambridge: Cambridge UP.

Loar, Brian. (1991) Mind and Meaning. Cambridge: Cambridge UP.

Loewer, Barry. (2001) "From Physics to Physicalism." in Physicalism and Its Discontents. Ed. Carl Gillett and Barry Loewer. Cambridge: Cambridge UP. 37-56.

Lowe, E.J. (2000) "Causal Closure Principles and Emergentism." Philosophy. 75: 571-585.

Lycan, William G. (1987) Consciousness. Cambridge, MA: MIT Press.

MacDonald, Graham, and Cynthia MacDonald. (1986) "Mental Causes and the Explanation of Action." Philosophical Quarterly. 36: 145-158.

Machamer, Peter, et al. (2000) "Thinking about Mechanisms." Philosophy of Science. 67(1): 1-25.

Mackie, J.L. (1974) The Cement of the Universe. Oxford: Clarendon.

Marras, Ausonio. (1998) "Kim's Principle of Explanatory Exclusion." Australasian Journal of Philosophy. 76(3): 439-451.

Matthen, Mohan and André Ariew. (2002) "Two Ways of Thinking About Fitness and Natural Selection." Journal of Philosophy. 99: 55-83.

McLaughlin, Brian P. (1992) "The Rise and Fall of British Emergentism." in Emergence or Reduction? Ed. Hans Flohr Ansgar Beckermann, and Jaegwon Kim. Berlin: W. de Gruyter.

---. (1993) "On Davidson's Response to the Charge of Epiphenomenalism." in Mental Causation. Ed. John and Alfred Mele Heil. Oxford: Clarendon. 27-40.

Meehl, P.E. and Wilfrid Sellars. (1956) "The Concept of Emergence." in Minnesota Studies in the Philosophy of Science. Ed. Herbert Feigl and Michael Scriven. Vol. 1, The Foundations of Science and the Concepts of Psychology and Psychoanalysis. 239-252.

Melnyk, Andrew. (1996) "Searle's Abstract Argument Against Strong AI." Synthese. 108(3): 391-419.

---. (1997) "How to Keep the `Physical' in Physicalism." Journal of Philosophy. 94(12): 622-637.

---. (2003) A Physicalist Manifesto: Thoroughly Modern Materialism. Cambridge: Cambridge UP.

Menzies, Peter. (1988) "Against Causal Reductionism." Mind. 97(388): 551-574.

Merricks, Trenton. (2001) Objects and Persons. Oxford: Clarendon.

Meyer, Richard A., et al. (1994) "Peripheral Neural Mechanisms of Nociception." in Textbook of Pain. Ed. Patrick D. Wall and Ronald Melzack. Edinburgh: Churchill Livingstone.

Millikan, Ruth Garrett. (1999) "Historical Kinds and the 'Special Sciences'." Philosophical Studies. 95: 45-65.

Montero, Barbara. (2003) "Varieties of Causal Closure." in Physicalism and Mental Causation. Ed. Sven Walter and Heinz-Dieter Heckmann. Exeter: Imprint Academic. 173-187.

---. (2006) "Physicalism in an Infinitely Decomposable World." Erkenntnis. 64: 177-191.

Moore, G.E. (1903) Principia Ethica. Cambridge: Cambridge UP.

Nagel, Ernest. (1961) The Structure of Science. New York: Harcourt, Brace & World.

Newell, A. and H. A. Simon. (1976) "Computer Science as Empricial Inquiry." Communications of the ACM. 19: 113-126.

Noordhof, Paul. (2003) "Not Old ... But Not That New Either: Explicability, Emergence, and the Charaterisation of Materialism." in Physicalism and Mental Causation. Ed. Sven Walter and Heinz-Dieter Heckmann. Exeter: Imprint Academic. 85-108.

Papineau, David. (1993) Philosophical Naturalism. Oxford: Blackwell.

---. (1995) "Arguments for Supervenience and Physical Realization." in Supervenience: New Essays. Ed. Elias E. Savellos and Ümit D. Yalçin. Cambridge: Cambridge UP. 226-243.

---. (1996) "A Universe of Zombies? The Problem of Consciousness and the Temptations of Dualism." Times Literary Supplement. 21 June: 3-4.

---. (2001) "The Rise of Physicalism." in Physicalism and Its Discontents. Ed. Carl Gillett and Barry Loewer. Cambridge. 3-36.

---. (2002a) Philosophical Naturalism. (Online revision of Papineau (1993)) http://www.kcl.ac.uk/ip/davidpapineau/Staff/Papineau/PhilNat2nded/PhNatIndexrevised.htm. Accessed: Oct. 12, 2005.

---. (2002b) Thinking About Consciousness. Oxford: Clarendon.

Peacocke, Christopher. (1979) Holistic Explanations. Oxford: Clarendon.

Pearl, Judea. (2000) Causality: Models, Reasoning, and Inference. Cambridge: Cambridge UP.

Pessoa, Luiz, et al. (1998) "Finding Out About Filling-In: A Guide to Perceptual Completion for Visual Science and the Philosophy of Perception." Behavioral and Brain Sciences. 21: 723-802.

"Physical". (1989) Oxford English Dictionary.

Piccinini, Gualtiero. (2004) "Functionalism, Computationalism, and Mental States." Studies in the History and Philosophy of Science. 35A(4): 811-833.

Pineda, David. (2001) "Functionalism and Nonreductive Physicalism." Theoria. 16: 43-63.

Poland, Jeffrey. (1994) Physicalism. Oxford: Oxford UP.

Polger, Thomas W. (2004) Natural Minds. Cambridge, MA: MIT Press.

Prior, Arthur N. (1949) "Determinables, Determinates and Determinants, Part I." Mind. 58: 1-20.

Putnam, Hilary. (1960) "Minds and Machines." in Dimensions of Mind. Ed. Sidney Hook. New York: NYU Press. (Reprinted in Putnam (1975a) pp. 362-385.)

---. (1963) "Brains and Behavior." in Analytical Philosophy Second Series. Ed. R. Butler. Oxford: Oxford UP. (Reprinted in Putnam (1975a) pp. 325-341.)

---. (1964) "Robots: Machines or Artificially Created Life?" Journal of Philosophy. 61: 668-691. (Reprinted in Putnam (1975a) pp. 386-407.)

---. (1967a) "The Mental Life of Some Machines." in Intentionality, Minds, and Perception. Ed. H. Castaneda. Detroit: Wayne State UP. (Reprinted in Putnam (1975a) pp. 408-428.)

---. (1967b) "The Nature of Mental States." ("Psychological Predicates") in Art, Mind, and Religion. Ed. Capitan and Merrill. Pittsburgh: U. of Pittsburgh Press. (Reprinted in Putnam (1975a) pp. 429-440.)

---. (1975a) Mind, Language and Reality: Philosophical Papers, Volume 2. Cambridge: Cambridge UP.

---. (1975b) "Philosophy and Our Mental Life." in Mind, Language and Reality: Philosophical Papers, Vol. 2. Cambridge: Cambridge UP. 291-303.

---. (1982) "Why There Isn't a Ready-Made World." Synthese. 51: 141-168.

---. (1988) Representation and Reality. Cambridge, MA: MIT Press.

Rees, Martin J. (2003) "Introduction." Philosophical Transactions of the Royal Society of London A. 361: 2427-2434.

Robb, David. (1997) "The Properties of Mental Causation." Philosophical Quarterly. 47: 178-194.

Salmon, Wesley C. (1984) Scientific Explanation and the Causal Structure of the World. Princeton: Princeton UP.

---. (1994) "Causality without Counterfactuals." Philosophy of Science. 61(2): 297-312.

Schaffer, Jonathan. (2003a) "Is There a Fundamental Level?" Noûs. 37(3): 498-517.

---. (2003b) "Overdetermining Causes." Philosophical Studies. 114: 23-45.

---. (2004) "Two Conceptions of Sparse Properties." Pacific Philosophical Quarterly. 85: 91-102.

Scheutz, Matthias. (1999) "When Physical Systems Realize Functions..." Minds and Machines. 9: 161-196.

---. (2001) "Computational Versus Causal Complexity." Minds and Machines. 11: 543-566.

Schiffer, Stephen R. (1987) Remnants of Meaning. Cambridge, MA: MIT Press.

Schröder, Jürgen. (2002) "The Supervenience Argument and the Generalization Problem." Erkenntnis. 56: 319-328.

Searle, John R. (1992) The Rediscovery of Mind. Cambridge, MA: MIT Press.

Shapiro, Lawrence A. (2000) "Multiple Realizations." Journal of Philosophy. 97(12): 635-654.

---. (2004) The Mind Incarnate. Cambridge, MA: MIT Press.

Shoemaker, Sydney. (1979) "Identity, Properties and Causality." Midwest Studies in Philosophy. 4. (Reprinted in Shoemaker (2003a) pp. 234-260.)

---. (1980) "Causality and Properties." in Time and Cause. Ed. Peter van Inwagen. Dordrecht. 109-135. (Reprinted in Shoemaker (2003a) pp. 206-233.)

---. (1981) "Some Varieties of Functionalism." Philosophical Topics. 12(1): 83-118. (Reprinted in Shoemaker (2003a) pp. 261-286.)

---. (2001) "Realization and Mental Causation." in Physicalism and Its Discontents. Ed. Carl Gillett and Barry Loewer. Cambridge: Cambridge UP. 74-98.

---. (2003a) Identity, Cause and Mind: Expanded Edition. Oxford: Clarendon.

---. (2003b) "Realization, Micro-Realization and Coincidence." Philosophy and Phenomenologial Research. 67(1): 1-22.

Sider, Theodore. (2003) "What's So Bad about Overdetermination?" Philosophy and Phenomenologial Research. 67(3): 719-726.

Sidtis, John J., et al. (1999) "Are Brain Functions Really Additive?" Neuroimage. 9: 490-496.

Smart, J.J.C. (1959) "Sensations and Brain Processes." Philosophical Review. 68(2): 141-156.

Smith, Peter. (1992) "Modest Reductions and the Unity of Science." in Reduction, Explanation, and Realism. Ed. David Charles and Kathleen Lennon. Oxford: Oxford UP. 19-43.

Sober, Elliott. (1985) "Panglossian Functionalism and the Philosophy of Mind." Synthese. 64: 165-194.

---. (1999) "The Multiple Realizability Argument Against Reductionism." Philosophy of Science. 66: 542-564.

Spurrett, David and David Papineau. (1999) "A Note on the Completeness of 'Physics'." Analysis. 59(1): 25-29.

Stalnaker, Robert. (1996) "Varieties of Supervenience." Philosophical Perspectives. 10: 221-241.

Steward, Helen. (1997) Ontology of Mind: Events, Processes, and States. Oxford: Clarendon.

Sturgeon, Scott. (1998) "Physicalism and Overdetermination." Mind. 107(426): 411-432.

Tabery, James G. (2004) "Synthesizing Activities and Interactions in the Concept of a Mechanism." Philosophy of Science. 71: 1-15.

Thomasson, Amie. (1998) "A Nonreductivist Solution to Mental Causation." Philosophical Studies. 89: 181-195.

Turing, Alan. (2004) The Essential Turing. Oxford: Clarendon.

Van Gulick, Robert. (1992) "Three Bad Arguments for Intentional Property Epiphenomenalism." Erkenntnis. 36(6): 311-331.

Wall, Patrick D. and Mervyn Jones. (1992) Defeating Pain. New York: Plenum.

Watkins, Michael. (2002) Rediscovering Colors: A Study in Pollyanna Realism. Dordrecht: Kluwer.

Wilson, Jessica. (1999) "How Superduper Does a Physicalist Supervenience Need to Be?" Philosophical Quarterly. 49(194): 33-52.

---. (2001) Physicalism, Emergentism, and Fundamental Forces. Ph.D. Dissertation. Ithaca, NY. Cornell University.

---. (2005) "Supervenience Based Formulations of Physicalism." Noûs. 39(3): 426-459.

Woodward, James. (2002) "What Is a Mechanism? A Counterfactual Account." Philosophy of Science. 69: S366-S377.

---. (2003) Making Things Happen: A Theory of Causal Explanation. Oxford: Oxford UP.

Worley, Sara. (1997) "Determination and Mental Causation." Erkenntnis. 46: 281-304.

Yablo, Stephen. (1992) "Mental Causation." Philosophical Review. 101(2): 245-280.