

**Pragmatism and Pluralism:
How to Avoid Becoming a Physicalist**

John Miller Ritchie

London School of Economics

PhD Philosophy

UMI Number: U615460

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

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



UMI U615460

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

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



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

THESES

F

7943



846653

Abstract

This thesis can be split into two sections. The first section is an argument against physicalism and its naturalistic pretensions. The second section presents an alternative to physicalism – pragmatic pluralism. The arguments for the first section are split into four chapters. In chapter one it is argued that an examination of science lends no support to physicalist ontology. The case of quantum chemistry is studied in some detail and it is shown that higher-level chemical facts are needed to support the quantum mechanical explanations given. The second and third chapters look at various ways in which physicalists have sought to explain away the apparent lack of unity in the sciences. Various accounts of the supervenience relation and functionalism are discussed and shown to be either inadequate for the physicalist programme or empirically implausible. The final chapter of this section discusses the so-called completeness of physics. It is shown that like physicalism in general, there is no way to formulate this doctrine to make it plausible in light of contemporary physics and able to underpin a physicalist ontology. The second section critically discusses alternatives to physicalism. The scientific pluralisms of John Dupre and Nancy Cartwright and the pragmatic pluralisms of Hilary Putnam and Nelson Goodman are discussed in detail. Drawing on the work of Putnam, in particular, and presenting Davidson's anomalous monism in a new guise, an original form of pragmatic or metaphysically deflationary pluralism is defended.

Table of Contents

Introduction: Physicalism: What Matters 4

Section 1: Against Physicalism

Chapter 1: Physics, Chemistry and the Science of Metaphysics 12

Chapter 2: Supervenience: Order out of Chaos 27

Chapter 3: Functional Analyses 68

Chapter 4: The Completeness of Physics 112

Section 2: Pluralism

Introduction to Section 2 147

Chapter 5: Science without Laws 150

Chapter 6: From Pragmatism to Pluralism 183

Chapter 7: Anomalous Pluralism 217

Conclusion: The Point of Pluralism 238

Introduction

Physicalism: What Matters

[T]he best guide we have to the nature of reality is the physical sciences.
David Armstrong¹

0.1 Something old, something new

What is the world like? That is a big question; and one might expect a correspondingly long and complicated answer, if any answer can be given at all. Philosophers, though, have traditionally provided one of four remarkably simple-sounding responses: materialism, idealism, dualism and, what I shall call for the lack of an agreed term, neutralism.

According to advocates of the first of these views, materialism, the world consists only of bits of matter, arranged in various ways. Among the great and the dead who might be counted materialists, we may include Democritus, Gassendi and Hobbes. Diametrically opposed to materialists are idealists. They say that the world is made up of nothing but ideas and minds. Berkeley, Kant and Hegel in different ways represent this world-view. Then there are the two halfway houses – dualism and neutralism. Dualists think that there are both bits of matter (like the materialist) and minds and ideas (like the idealist) but these two sorts of entity are fundamentally different, neither can be reduced to the other. Descartes is the father of all such theories. Neutralists think that the fundamental stuff of the world is neither matter nor mind or at least not solely matter or solely mind but something else which gives rise to matter and mind. Spinoza argued for something like this, as did Russell once. They have modern disciples in Thomas Nagel and David Chalmers.

¹ Moser and Trout (eds.) (1995), p.45

As ever in philosophy, traditional answers have traditional problems. How can materialists explain the workings of the human mind, the mysteries of consciousness and intentionality, in terms of the redistribution of matter? How can idealists explain the fact that the world does not seem to be made up of ideas but of people and cars and biscuits and oysters and phone boxes and sand and all kinds of other stuff? How can dualists explain the connection between matter and mind? What on earth is this stuff that neutralists are talking about? These sorts of question have broadly speaking set the terms of the debate since the time of Descartes – at least as I say for metaphysicians.

A non-philosophical way to answer the same question – what is the world like – would be just to list the sorts of things that we come across there. A better way might be to try to create some sort of system to describe these objects and their interactions. Arguably, this is what the sciences (or at least some of them) try to do. By extension then, the sciences also address the question ‘what is the world like?’. How should we understand the relation between these two projects – the scientific and the philosophical?

The most popular contemporary answer sees them as complementary. Materialism has been recast as physicalism; a doctrine which sees itself in broad outline as a successor to the views of Democritus, Hobbes and others but which derives its detail from contemporary science – in particular physics. There are at least two reasons for such a change: one to do with the content of materialism, the other to do with its justification. First, it is clear that although old-style materialists were in some sense right that there exists tiny atoms, the way they conceived of them was seriously mistaken. What we now know about atomic structure suggests that atoms are far from the tiny billiard-ball-like objects that

many materialists had in mind. Moreover, physicists have come up with all sorts of other entities, fields for example, which seem to have no counterpart in past materialist metaphysics. So modern physicalism has shifted away from the atoms-in-the-void conception of the world to one in which they defer to physicists for information concerning the fundamental constituents of the universe. Modern physicalists, in other words, want to stop *a priori* speculation about how the world is and, as David Lewis (1986) puts it, “take sides with physics”.

More significant than this change of content is the belief that physicalism is the ontology employed in and supported by the sciences.

Broadly empirical character [physicalism is] supported inductively by scientific practice. (Hellmann and Thompson (1977))

According to contemporary physicalists, the principles of physicalism are to be treated as high level empirical generalisations. (Post (1991))

On the theory side, developments in science, successful reductions of one discipline to another for example, are supposed to confirm the physicalist ontology. In terms of practice, it is often argued that physicalism is presupposed in much scientific inquiry and this leads to fruitful research. So there is a descriptive and a normative part to the argument from science. On the descriptive side, physicalists will claim that the way science portrays the world and the way scientific descriptions seem to be developing lends support to physicalism. On the normative side, presupposing physicalism, it is claimed, makes for good science.

Even when physicalists seem to be offering justifications that rely on other considerations, it becomes clear that whatever force such arguments have depends on the implicit understanding that physicalism has the backing of the sciences. For example, one common argument heard in favour of physicalism is that it is the simplest explanation of how the world is. Setting aside the question of why simplicity should be seen as a guide to truth, the argument, unless fleshed out, cannot discriminate between physicalism, idealism and neutralism. It only has any force if we already accept the idea that science correctly describes most of the world in physical terms. Then the choice is simply between dualism and physicalism – and physicalism is clearly simpler.

Another argument sometimes put forward is that any metaphysics other than physicalism involves believing in some kind of ‘spooky’ ontology – weird mind-type substances and the like. But again this by itself is a woeful argument. What one does and does not consider ‘spooky’ tends to be a function of one’s favoured ontology. For example, Berkeley given his theory of ideas found the notion of matter incoherent. Again, if this argument has any force at all, it is only given the presupposition that natural science backs up the idea that the world is largely as physicalists take it to be. Therefore to advocate any other ontology is to go against the sciences, obviously a bad thing. Arguments for physicalism cannot be divorced from the idea that science supports this ontology and only this ontology.

Contemporary physicalism then is the combination of two ideas: materialism and what we might call naturalism. It retains the idea that the world is made up of some fundamental constituents but lets physicists tell you what that something is; and it seeks ultimate justification through the practice and theory of

the sciences. It is important to see that this is not merely the handing over of the question of how the world is to the scientific community. Physicalist philosophers want to take the claims of the physicist and, like their materialist forebears, expound a metaphysical thesis which goes beyond the current claims of the sciences. If we were to try to sum it up in a sentence, physicalists believe that everything is, or is dependent upon, the physical.

There remain some ambiguity in the definition. One needs to clarify what is meant by the physical in physicalism. Narrowly construed, we might take it to refer only to the discipline of physics itself, so the science which fills in the content of old-style materialism is only physics. More broadly construed, physicalism could refer to all the physical sciences, so that would include chemistry and maybe biochemistry as well as physics. Many physicalist seem to take their doctrine in the broad sense only. Hence they consider the only explanatory tasks for the physicalist are to fit the life sciences and the mental into their favoured metaphysical picture.² One possible reason for such a liberal attitude might be the thought that the relation between the other 'physical sciences' and physics is unproblematic. Perhaps it is even thought that successful reductionist treatment of these domains provides data to support the claim that physicalism is confirmed by science. Whatever the reason, once we recognise that there are domains that are not covered by physics, other than just the mental or the biological, the case for physicalism must be made there too. If physicalism is genuinely meant to be a naturalistic metaphysics, it will have to prove itself in all domains.

² See for example David Papineau (1993).

The thesis divides into two sections. The first section is an argument against physicalism. The second section is an attempt to provide a better answer to that question.

The first chapter of this thesis will discuss the support science lends to physicalism and in particular will discuss the relation between physics and chemistry in the context of quantum chemistry. The sketch I provide of that relation will form the basis of a claim that I will make regarding the state of science in general, viz., no neatly hierarchical or reductionist ontological conclusions can be drawn from the current state of the sciences. I do not consider this claim either surprising or in itself of great philosophical interest. It is, I take it, a fact that anyone acquainted with the practices of the sciences will be familiar with. Sociological evidence for this can be found in the oft-repeated idea that even scientists working in similar fields find it difficult to communicate their theories and ideas to one another. The messy and complicated relations of the sciences form the background against which all metaphysical discussions that purport to be remotely naturalistic must take place.

Given this background, two questions arise: how the physicalist should articulate his ontology and more importantly why physicalism is the right ontology to be articulating in the first place. The second and third chapter of this work will deal with the first of these points: two sorts of ways of connecting the physical to the non-physical that supposedly do not commit one to reductionism (and therefore allow for a certain conceptual messiness) will be discussed – supervenience and functionalism. The fourth and last chapter of the first section will discuss an argument for physicalism that does not rely on articulating precisely the relation between physics and the other sciences: the argument from

the completeness of physics. The thrust of this argument, as I interpret it, is that certain widely accepted facts about physics itself oblige us to be physicalists. Again the dialectic strategy will be to relate this claim to actual scientific theories and show that if physicalists have anything like the sort of physics employed by scientists in mind, this argument is a non-starter.

I hope to convince the reader that there is little or no support given by the sciences for the particular ontological commitments of physicalists and that the so-called arguments for physicalism, whether they be descriptive or normative in character, are either empty and rhetorical or simply invalid. In short, physicalism, it will be claimed, is no more a naturalistic metaphysics than any of the alternatives.

0.2 The rights and wrongs of physicalist metaphysics

Physicalism, in intent, if not implementation, contains a thought worth pursuing. The idea that we should move away from traditional metaphysical assumptions and positions and look to the sciences to tell us how the world is, is one I applaud. The problem with contemporary physicalism, I shall argue, is that it is still bound by some of the claims of older materialisms into exaggerated, unwarranted or empty claims. The second half of this thesis will be an attempt to liberate our thought about how the world is from the traditional metaphysical positions listed above; an attempt to remove the metaphysical backdrop and articulate a pragmatic pluralism. Not only is this more in tune with the broadly naturalist objectives of the physicalist programme, but I shall argue that it is also inevitably the only position that we can make sense of. We are all *de facto* pragmatic pluralists.

Section 1
Against Physicalism

Chapter 1

Physics, Chemistry and the Science of Metaphysics

1.0 What sort of evidence?

In what way are the disciplines of physics and the other sciences meant to lend support to the metaphysics of physicalism? There are at least two arguments or kinds of argument that are used. First, it is often claimed that general considerations about the subject matter of physics itself and its supposed universality force physicalism upon the philosopher on pain of absurdity. Arguments from the so-called completeness of physics fall into this category. (They will be discussed in detail in chapter 4.) Of present concern here will be another argument. An argument that seeks justifications of physicalism that proceed not from consideration of physics itself but the relation physics bears to other disciplines. What I shall call arguments from reduction.¹

1.1 Arguments from reduction

Arguments from successful reduction might be thought to work in two ways. Either as direct evidence that all there is to some area can be captured by physics

¹ Another argument that has sometimes been put to me in person but that I cannot find anywhere in the literature is that cosmology forces us to be physicalists. The argument can again be given in descriptive or prescriptive form. On the descriptive side, it might be suggested that what we already know about the beginning of the universe implies a physicalist ontology. Since the world had a physical beginning, it might be argued that it cannot be anything other than result of basic physical interactions, unless you are willing to posit some novel and weird sorts of interaction or law to explain the existence and workings of the apparently non-physical. On the normative side, one might contend that if a science like cosmology is possible, it must be premised on the idea that the complete evolution of the universe can be explained in terms of the elements that occurred at the beginning of the universe. In other words, since what occurred at the beginning was physical, everything must be explicable in terms of physics. The fact that such arguments rarely surface is I think tacit acknowledgement of the fact that cosmology is not the ideal science on which to base physicalism. Cosmological theories are so speculative and so highly susceptible to revision it would be a brave individual who based his metaphysics on such results. Also phenomenon like spontaneous symmetry breaking make it far from obvious that cosmologists do behave as the argument-sketch I have provided suggests. See Auyang (2000) for a discussion of symmetry breaking.

or as part of a more general meta-inductive argument. Clearly single cases of reduction can lend no *general* support to physicalism; such reductions show only one particular part of the world is captured by physics. Thus, without a complete reduction of all disciplines to physics (which I take it no one believes we have) the only argument that can proceed from reduction is the meta-inductive one.

Putting it roughly (for that is the only way it has ever been put) the success of physics in providing reductions in (say) chemistry and biochemistry leads one to expect similar successes for other disciplines, most notably psychology. Generally those attracted to reductionism assume that there is a hierarchical structure to the world²; starting at the bottom with physics and progressing through chemistry and biology eventually to psychology and maybe sociology. Each level is supposed to represent a new level of complexity that can be explained by the level below and ultimately, in theory at least, physics. As it stands this picture is no doubt far too simple: much of what is rightly termed physics involves complex entities – galaxies for example. Nevertheless, this general hierarchical picture pervades and informs much physicalist-reductionist thinking. The argument from reduction thus proceeds from successful cases of reduction to the conclusion that it is reasonable to believe that all domains within the hierarchy must be reductively related to the domain below; and hence, all domains are ultimately reducible to physics.

There are a number of issues that need to be discussed here, not least what we understand by the term reduction. However, first I want to draw the reader's attention to what seem to me basic problems with an argument of this form. Granting for the moment that there do in fact exist such reductions, the meta-

² See Oppenheim and Putnam (1958) for a classic account.

inductive argument, as stated here, strikes me as unsustainable for two reasons. First, it is not at all clear why success in one area should lead one to expect success in another. At the very least, some argument is needed to justify the idea that all other, as yet unreduced domains can be reasonably be expected to conform to the pattern of the supposedly successful reductions. In other words, that we can, in fact, perform a reliable induction over such cases. After all, it might well be part of the traditional dualist picture to expect physics to explain everything in just the way physicalists envisage, *except* for the mental. So there needs to be some justification of the idea that reductionist principles will extend into every domain.³ Second, if this meta-inductive argument is to be successful, as with any inductive argument, the evidence must all point in the same direction. Physicalists must expect that these reductive successes point towards the same underlying physical ontology. But that just does not appear to be case. Consider two plausible candidates for reduction: thermodynamics and the chemical concept of valence. It might be argued that thermodynamics reduces to statistical mechanics⁴ and that valency can be explained in terms of a quantum mechanical treatment of electron orbitals. So for the physicalist these two ‘successes’ might be considered data points in their inductive argument. However, as far as the ontology is concerned (and that is what is at issue here) they provide no reliable pattern at all. The ontology of statistical mechanics is radically different from that of quantum mechanics. One employs a notion of particles with definite position and trajectories and the other does not. One theory uses a classical account of

³ As Sober (1999) remarks the meta-inductive argument can make it sound as though reductionist successes were balls drawn at random from an urn. Clearly, though, scientists have not achieved their reductionist successes by random sampling from all of science. They have investigated areas in which they believed such reductionist ploys were likely to succeed, given the techniques and tools that they had available to them.

probability; the other does not. One assumes an underlying deterministic ontology; the other does not. The differences are many and varied. There just is no general ontological picture that can be inferred from these two cases, even if we grant that they are both successful reductions. At the very best, such cases point towards the *methodological* maxim that searching for underlying mechanisms can often be fruitful. But this principle, as I shall argue below and in subsequent chapters, has a life quite independently of the ontological claims of physicalism.

I believe such considerations count powerfully against arguments from reduction. However, I now intend to shift tack and argue that that the general presupposition that underpins the meta-inductive argument is flawed: namely that there exist many clear-cut cases of reduction which can form the premises of the meta-induction. As I mentioned at the outset many physicalists believe that reductions of chemistry to physics in particular provide powerful reasons to believe there should be reductions elsewhere. By looking at quantum chemistry I shall show that this simple reductionist story misrepresents the relation between modern physics and chemistry.

1.2 Reduction: formal and informal

Before we can assess the plausibility of the claim that reductions in science provide some kind of evidence for physicalism, in particular reductions of physics to chemistry, it is necessary to have some idea of what a reduction is. Traditional formal accounts have focused on the idea of deducibility of one theory from another. According to Ernest Nagel's classic account, a reduction is effected using bridge principles (generally biconditional statements) connecting the terms of one

⁴ This claim is of course often challenged. See for example, Sklar (1995) for an account of the

theory to another. The reduced theory is then deduced via the bridge laws from the reducing theory. So for example there is a bridge law connecting the term temperature in thermodynamics with mean molecular kinetic energy in statistical mechanics. Using such a bridge principle, Nagel claims we can deduce thermodynamics (or some part of it) from statistical mechanics.

Nagel's account has largely fallen out of favour. Doubt has been expressed about whether bridge laws provide the right sort of connection between the terms (or the properties) of the reducing and the reduced theory. In particular, one might wonder whether a number of relations that intuitively speaking are not of the right sort might count as bridge laws; for example, properties connected by causation. More importantly, though, the bridge laws themselves demand some kind of explanation. Questions arise concerning their ontological status: are they on a par with the laws being reduced? If so, then the very idea of reduction seems to be undermined, since there are true laws which ineliminably involve the properties of the supposedly reduced theory. If not, then precisely what is supposed to differentiate them from other laws? There is also the danger of trivialisation. If there are no restrictions on the kind of statement that counts as a bridge law, then it is possible to deduce one theory from any other simply by introducing a bridge law which states that if the reducing theory holds then so does the reduced theory. That is to say take our reducing theory to be A (say Newtonian mechanics), our reduced theory B (say, that Englishmen are bad losers), B can be deduced from A, given the bridge law $A \leftrightarrow B$.

Robert Causey (1980) offered a natural strengthening of the Nagel model by replacing bridge laws by identity statements.⁵ The benefit of this approach is that identity statements require no further explanation, metaphysically speaking at least. If water *is* actually H₂O, then one does not need to explain why wherever there is water there is H₂O. (Although, of course, one would expect there to be some justification for positing the identity statement in the first place.) Moreover, identity statements cannot be arbitrarily invented. Philosophically speaking this is certainly an improvement. Causey's account clearly does away with the need to explain the bridge law. However, it seems to rule out almost all putative cases of reduction from the history of science. It is difficult to find even one example of a deduction of one theory from another employing only identity statements to connect the properties of the different theories.

Consider the example of light and its supposed reduction to electromagnetic radiation. On a standard and simplified account of this episode from the history of science we might say that what is termed 'light' in one theory is identified with electromagnetic radiation in the other. But this ignores the ontological disparities between the then extant theory of light and electromagnetism. Fresnel's theory of light, the most up-to-date before the advent of electromagnetism, described light as propagating through a solid, mechanical elastic medium (an ether). In Maxwell's theory light is described as a periodic disturbance in an electromagnetic field. As John Worrall (1988) has said "[o]ne would be hard pressed to cite two things more different than a displacement current, which is what [electromagnetism] makes light, and an elastic vibration

⁵ Arguably this is also Nagel's view. (The issue is somewhat clouded by whether or not you adopt a nominalist line with regard to properties and therefore shy away from property identities. A fuller investigation of Nagel's views would require more discussion of this point. However, such niceties of exegesis shall not concern us here.)

through a medium, which is what Fresnel's theory [makes it]." So the identification of light (as understood by Fresnel) with electromagnetic radiation is highly problematic.

Acknowledgement of this kind of radical ontological change opens up another possibility for the physicalist. Instead of claiming that the relation between physics and other disciplines is best described in reductionist terms, she may think it is better to take an eliminativist line. That is to say, non-physical theory is to be taken to be largely false and at best an approximation to the true physical description of the world. Such a picture undercuts the hierarchical image of the sciences described above. Given an eliminativist conception, there would be a hierarchy of descriptions, most of which are false, but no hierarchy of reality. We shall have occasion to examine eliminativism and some of its consequences in later chapters. One should merely pause to note here that there is no eliminativist analogue to the reductionist argument for physicalism. The eliminativist will agree with the non-physicalist that certain theories cannot be reduced to physics. However, the eliminativist will draw the conclusion that such theories are therefore false; the non-physicalist, on the other hand, will conclude rather that physicalism is false. One can only arbitrate between such positions by considering the arguments for physicalism that rest upon physics itself. For this we will have to wait until chapter 4.

In any case, as perhaps my example suggests, the distinction between elimination and reduction is a matter of degree, rather than a hard and fast one. There are some cases which seem more plausibly reductions, others perhaps that are better described as eliminations. More sophisticated accounts of reduction have been developed by others, notably by C.A. Hooker (1981) and Paul

Churchland (1979), to deal with such problems. However, here I am not much concerned with the technical details of reduction and the semantic and epistemological difficulties they create. All we require is a definition that will allow us to assess the plausibility of the meta-inductive argument; and as others have argued a rough and ready idea of the notion can easily be given, even if we lack the highly refined analysis that some philosophers seek. Peter Smith (1992) nicely sums up this rough and ready reductionist account:

[T]he essence of reduction is simply an explanation in terms of one theory of why another [theory] works and [such] explanations come in a variety of flavours (p.21)⁶

Let us apply this informal notion of reduction to the case of quantum chemistry.

1.3 Reductionism: a watery grave?

According to the Smith-style basic notion of reduction, if chemistry is reducible to physics then we should expect an explanation of chemical phenomena in some fashion or other purely in terms of physics. Let us consider quantum chemical explanations of some of the spectroscopic properties of water.

If one is of a reductionist turn of mind, one would expect the quantum chemist to write down a Schrödinger equation (the equation which describes all interactions in quantum mechanics) solely in terms of fundamental physical interactions for the molecule in question. However, this is not what happens. What quantum chemists seek is a model in which they can apply their theory and

provide some useful and tractable results. Here is a brief qualitative summary of how one might go about investigating the spectroscopic properties of water molecules:

Spectroscopic properties of any molecule (or atom) are determined by the energy levels of the electrons. As an electron falls from one energy level to another, a photon with a frequency corresponding to the energy difference between the two levels is emitted. The quantum chemist thus wants to come up with some way of calculating the value of these energy levels. To do this she has to be able to model the sorts of interactions that dictate the features of the Hamiltonian, i.e., the part of the Schrödinger equation that describes the possible energy functions of the orbital electrons. A basic construction of the Hamiltonian might begin something like this. First one would assess the sorts of interaction a molecule like water might undergo. It is known from chemistry that the water molecule is a three atom, non-linear structure with one heavy nucleus bonded with two lighter nuclei. One would therefore expect there to be vibrational motion between the nuclei, along the bonds, as well as rotational motion for the structure as a whole. The quantum chemist can model these vibrational modes using the harmonic oscillator (with a few adjustments); in other words by treating the movement between the bonds like the quantum version of a spring. For the rotational modes a rigid rotator model is used, which is the quantum analogue of the classical motion of rotation without deformation. One can solve these Hamiltonians and obtain theoretical values for the electron energy levels, which in turn can be used to calculate the value of the spectral lines. More accurate results can be obtained by adding perturbations – for example coupling effects, and

⁶ Note of course that this is arguably what Nagel's model is a version of, employing the deductive-

possible deformations to the basic molecular structure. Alternatively, so called variational calculations may be made. Essentially, this involves exploiting certain mathematical properties⁷ of quantum mechanics to obtain an approximate solution to the ground-state energy for the molecule by varying certain judiciously chosen parameters. Here again knowledge of the structural properties of water is essential in applying this strategy as well as modelling the Hamiltonian. The sort of co-ordinates that one chooses to model the system is dictated by knowledge of the structure of water.⁸ Moreover, in making the actual calculations when employing this method, one has to know certain facts about the typical bondlengths and angles of water in order to break it down into computationally tractable chunks. So what we see, in fact, is that prior knowledge of classical chemistry is imported into the quantum chemist's treatment at every stage. First, in modelling the kind of motions that one would expect the molecule to undergo and thus allowing the theoretician to write down appropriate Hamiltonians. And second, knowledge of the typical values for certain features of the molecule are necessary before one can even begin to make calculations. Generally, one can say that the application of quantum theory depends on prior knowledge of the structure of the water molecule.

What we find then is that quantum chemical explanation of spectroscopic effects must assume certain facts about the structure of the molecule, facts which are not derived from fundamental theory but from classical chemistry. Now, as

nomological model of explanation.

⁷ The exact energy values for any Hamiltonian are known as eigenvalues, and corresponding to each there is a wave function, an eigenfunction. Any arbitrary wavefunction can be written as the sum of these eigenfunctions. One can then show that the expectation value for the function must be greater than or equal to the ground state energy level. The approximate ground state is thus found by varying (hence the name) the parameters of the arbitrarily selected function until one obtains the lowest possible value. My thanks to Sophie Kain for her careful explanation of this and quantum chemistry in general.

⁸ Radau co-ordinates. Those interested in the technical detail are referred to S. Kain's thesis

said at the outset, I do not expect anyone to be particularly surprised by this description of quantum chemistry. A physicalist could argue that although there is in fact no reduction here, that does not mean there could not be a reduction *in principle*. It is, after all, by physicalists' lights, not mysterious why one cannot explain the spectroscopic effects in purely physical terms. Schrödinger equations which attempted to describe only basic physical, in particular Coulombic, interactions for this system would be mathematically intractable. Given that the direct route to reduction is blocked off, the only way one could explain the spectroscopic features of water is by introducing knowledge from other domains. I have no objection to offer at this point⁹ to that response but one should bear in mind that it reverses the order of the explanation. Instead of arguing *for* physicalism, one is now justifying the apparent failures of reduction *assuming* physicalism. In short, science is no longer supporting physicalism, physicalism is explaining away certain features of science. This is, of course, a game any metaphysical theory can play and physicalism's claim to be special, to be a naturalistic metaphysics, is thus weakened.

The above considerations provide yet another nail in the already well-sealed coffin of the meta-inductive argument. There are, in reality, very few data-points for on which to base the meta-inductive argument. As the above discussion shows, even favourite cases like chemistry turn out to be problematic for physicalists and there are many other, more detailed case studies that demonstrate the same point for other disciplines. For example, Nancy Cartwright (1999) and Sunny Auyang (2000) have independently discussed cases where reductions in different areas of physics are seen to be problematic; Sahorta Sarkar (2000) shows

⁹ However, in chapter 4 I will argue that this explanation is unsatisfactory for other reasons.

the inadequacy of thinking that molecular biology reduces to physics or even chemistry; John Dupré (1993) discusses similar cases in biology more generally. Moreover, the whole philosophical world knows the enormous number of anti-reductionist arguments that have been put forward regarding the mental. The conclusions the various authors wish to draw from such cases are as varied as the areas they have chosen to discuss. What is undeniable though is that their case studies, taken together, show that the world described by the sciences is a mess. A whole host of theories that make all kinds of different ontological assumptions are employed to describe the world. Contemporary science provides nothing like the sort of picture that could support the meta-inductive argument.

Physicalists might argue that such considerations are not of the greatest importance. “Of course,” they might say, “until we have the full reductionist story it will be easy to point out failures. However, if we look at the history of science, then one sees the long-term success of reductionism. It is this rather than current science, which physicalists wish to base their meta-inductive argument upon.” But the history of science is littered with as many, if not more, reductionist failures as successes: the attempt to reduce electromagnetic phenomena to mechanics; Einstein’s attempt to come up with a unified field theory, a whole host of discarded unification projects in particle physics. Again, there is an easy physicalist reply to such failures – they don’t show the inadequacy of physicalism *per se*, just of the particular physics being used in the reductions. But physicalist can hardly claim then (unless they know the one true physics in advance) that any part of science lends support to their metaphysical position. As before this kind of physicalist response would merely explain away science and its history, rather than draw from science support for physicalism. The history of science, as much

as contemporary science tells a tale of theoretical disunity.¹⁰ The data now and in the past is unequivocal; there is no curve of best fit converging in the limit to a physicalist ontology. This is a theme that will be elaborated in chapters 4 and 5.

We may summarise our discussion of the meta-inductive argument as follows. The argument in the abstract is flawed because it provides no basis for thinking that one can reasonably infer that successful reductions in one area will lead to successful reductions in another. Moreover, the putative cases of reduction provide no pattern from which a physicalist ontology might be inferred. Once we look at science in detail we see that even favoured cases of reduction turn out to be complex and at best partial; and the history of science testifies as much to the failure of reductionist programmes as to its successes.

Let me reiterate what I take it has been shown: there is no argument from reduction in science to physicalism. We should conclude no more than this from the arguments so far given. It certainly has not been shown that physicalism is false. All we are entitled to claim is that there is no support for physicalism from this quarter.

1.4 Methodology and metaphysics and the methodology of the metaphysician

Let me separate my anti-reductionist observations from possible anti-reductionist methodological conclusions one might want to draw. I don't want to suggest that scientists should not look for reductions or underlying causal mechanisms. Such a tactic has proved fruitful in the history of science even if it has not lead to completely successful reductions. But a commitment to such a methodological

¹⁰ I agree with Ian Hacking (1983) when he says: "Every single year...physics alone has used

principle in no way forces upon one any metaphysical position which has implications beyond the domain one is investigating. As I have suggested for different sciences, the idea of an underlying mechanism and what one hopes to explain with it will be different. There is no *global* metaphysics which follows from the sage advice that it often helps to understand how the parts work, if you want to get a better understanding of the whole. Even this concession has to be tempered with the observation that like all pieces of sage advice, it is limited in its application. Sometimes looking for underlying mechanisms is just not the right thing to do. For example, thermodynamics and the theory of relativity are derived from some general and empirically supported statements, not an investigation of underlying mechanisms. Sometimes this axiomatic approach proves fruitful but one cannot tell in advance; as bohemians say, you have to try these things. I shall have more to say on the role of metaphysics in methodology in chapter 7.

successfully more (incompatible) models of phenomena in its day to day business.”

What I have said thus far will probably meet little opposition from many contemporary physicalists. Seldom, today, do physicalists seek to justify their position by invoking arguments from reduction. Generally a more abstract, bipartite form of reasoning is employed. On the one hand, physicalists seek to explain how the non-reducibility of non-physical domains is compatible with a physicalist ontology. On the other hand, justifications for the physicalist doctrine are sought from general features of physics. This divides the tasks between the scientists and the metaphysicians equally. It is the metaphysician's role to explain to us how there can be ontological unity beneath phenomenological disunity. It is up to the scientist, more specifically the physicist, to provide evidence for the truth of the doctrine.

Some physicalists will perhaps not see the need to carry out the first part of this task. They will be content to rest with the idea of reduction *in principle*.¹¹ I think that is a mistake. Since we have no general characterisation of what reduction involves, this claim strikes me as empty.¹² The braver physicalists will attempt to articulate exactly what sort of relation might exist between physics and other disciplines which apparently do not reduce to it. It is to these braver souls I turn next.

¹¹ Eliminativists need only concern themselves with the latter task as well.

¹² John Worrall has suggested to me the idea that reductionist thinking is captured in the idea that we should try to do with as few primitives as possible. But this is surely just a platitude. Who has ever suggested we should have more primitives than are necessary? What is at issue is whether it is possible to capture everything using just physical primitives (whatever they are).

Chapter 2

Supervenience: Order out of Chaos

Science bequeaths the metaphysician an untidy world. Many physicalist philosophers, unsatisfied with this mess, take it upon themselves to uncover the order that lies deep within. Two ideas are used to effect this trick of finding order out of chaos: the completeness of physics and supervenience. The completeness of physics provides the order. It will be the topic of a later chapter. First, though, I wish to turn to the principle which explains the apparent chaos – supervenience. Supervenience, so it is hoped, will bind the physical to the non-physical in a way that will allow the metaphysician to retain the label physicalist without commitment to reductionism. The combination of these two views results in the metaphysical orthodoxy in contemporary philosophy, *non-reductive* physicalism. The non-reductive position shares many assumptions with the ordinary classical reductionist picture. It retains the idea that the world can be split into many levels: the physical, the chemical, the biological, etc. However, the notion that we should be able to reduce these to one base level is given up.

Historically at least, the main motivation for this retreat from reductionism has not been the phenomenology of the sciences, outlined in the previous chapter but what are called *multiple-realisation arguments*. Such arguments make the, perhaps obvious, point that certain kinds of property or state can have many physical bases. For example, the computational states required for the calculation $2+2=4$ can be realised by a silicon-based calculator, a carbon-based life-form and a host of other more or less complicated mechanisms. If a property or state can be multiply realised, then there can be no hope of identifying it with any one single,

unified physical base and hence reduction is blocked.¹ At least, that has been the general texture of the argument traditionally given² but recently doubt has been cast on the conventional view that multiple realisation blocks reduction, notably by Jaegwon Kim (1993, 1995). Whether or not this is the case will not concern me here greatly. What I want to investigate is the plausibility, the pros and cons of this orthodox position (whether or not one ultimately decides that it is reductionist or not). In short, the chapter is concerned with the plausibility of physicalistic views defined using supervenience. (For the sake of simplicity I shall refer to physicalist accounts employing the concept of supervenience as non-reductive throughout.)

The topic of supervenience has spawned a large technical literature so some definitions and clarifications are in order first. The following discussion will be abstract and I shall, at least to begin with, simply make reference to physical and non-physical properties without detailed discussion of what these might be. Mainly, in accordance with the literature, I will talk of the relation between physical and mental properties. That is to say, I shall assume whatever the basic physical properties are they do not include mental properties. Occasionally, though, I shall have cause to talk of the relation between physical and other non-physical properties. When I do so I shall make this clear. Later sections will tackle the issue of defining the physical in greater detail.

¹ Clearly this account interprets reduction in terms of something like the Causey-model in which bridge laws are identity statement.

2.1 Defining supervenience

The origin of the term ‘supervenience’ in the context of the modern mind-body problem³ can be traced to Donald Davidson’s seminal paper “Mental Events”. There the idea is presented somewhat informally as ‘no change in the mental without a change in the physical’.⁴ The technical explication of this basic idea is mostly, if not solely, due to the efforts of Jaegwon Kim. From his work several permutations of supervenience have arisen which are standardly classed into three different types of relation. One set of possible definitions runs as follows:

(SS) Strong Supervenience M strongly supervenes on P just in case necessarily for any object x and any property F in M, if x has F, then there exists a property G in P such that x has G, and necessarily if any y has G, it has F.

(WS) Weak supervenience M weakly supervenes on P just in case necessarily for any object x and any property F in M, if x has F, then there exists a property G in P such that x has G, and if any y has G, it has F.

(GS) Global Supervenience M globally supervenes on P just in case for any two worlds, w_1 and w_2 , if they are P-property indistinguishable, then they are M-property indistinguishable.⁵

² See Putnam (1960), Fodor (1974), Boyd (1980) and many others.

³ Certain emergentists also used the word in the first half of the 20th century.

⁴ In fact I will not discuss Davidson’s ideas in this chapter. The special context in which his ideas arise deserve separate and detailed attention (which they get in chapter 7).

⁵ Different philosophers will defend subtly different variations of these relations depending upon on how they construe the modal terms (e.g. whether a possible worlds analysis is appropriate or not) in these definitions and depending on whether they think the supervenience relation holds for objects or events. Those with a taste for such nuances are referred to Brian McLaughlin’s paper

Where M and P are non-empty families of properties.

To put the general idea in English: if the mental strongly supervenes on the physical then wherever there are certain physical properties there will be certain mental properties and that relation holds counterfactually. (WS) is identical, except the relation does not hold counterfactually. (GS) simply implies that given the complete physical state of the universe, the complete mental state of the universe is fixed. Intuitively (although there is room to disagree depending on how one construes the modal terms) (SS) implies both (WS) and (GS). Note that although the above definitions of (SS) and (WS) are framed in terms of relations between properties copresent in an individual, the definitions need not be that restrictive. One could easily rephrase both (SS) and (WS) in terms of a relation between properties in a certain region. That is to say, take some region of space. Then fix all the P-properties intrinsic to that area and (according to this form of supervenience) all the M-properties are fixed too. Terence Horgan calls this regional supervenience. I prefer to think of it as a more general definition of (WS) or (SS).⁶ In the limiting case where the subvening properties include all physical properties then (SS) becomes a version of (GS).

Let us turn to then to the question of the appropriateness or otherwise of these forms of supervenience as explanations of the mind-body relation. There is a powerful and well-known objection to (GS) which goes along the following lines. Assume that mental properties globally supervene on physical properties. Now, imagine two worlds which are physically identical and hence, by (GS), mentally

⁶“Varieties of Supervenience” in Savellos and Yalçın (1995). These nuances will not affect my criticisms.

identical. Let us suppose there is some minor change in physical properties in one of the worlds: that for example, a few hydrogen atoms are shifted in a galaxy far, far away. The two worlds now differ physically. It is perfectly consistent with (GS) that the two worlds may differ radically in the distribution of their mental properties (or that there may be no mental properties at all in one of the worlds). It seems therefore utterly mysterious what the relation between physical and mental properties is, given (GS) or why one should believe that the mental is actually *dependent* upon the physical in any interesting sense at all.⁷

In response, John Post (1995) has argued that although global supervenience is *consistent* with the sort of story told above, it does not *imply* its truth. It may well be that moving a few hydrogen atoms has no affect at all on the mental properties of the universe. Post suggests that (GS) be understood as programmatic. It defines the minimum structure which physicalism has to satisfy, the detail of which will be filled in later. However, when one considers what that extra detail might be it is difficult to resist the thought it will involve citing particular physical bases upon which particular mental states supervene. In other words, it will involve setting out the sorts of relations that characterise (SS); at least as liberally construed by me. I do not see any way of defending (GS) as providing an explanation for the link between the mental and the physical that does not make it into a version of (SS).

Weak and strong supervenience look more promising alternatives. They are clearly explanatorily more robust forms of the supervenience relation because they tie the supervenient property more closely to the physical manifestation of

⁶ Note that the above definition I have provided will already be an adequate account of regional supervenience, if one is liberal in what one counts as an individual and as property that may be predicated of it.

⁷ Kim (1989) first makes this point, as far as I am aware.

the thing we take to have that property. That is, if the mental supervenes on the physical it supervenes on the sort of physical thing that has mental properties, i.e. human bodies (or human bodies plus the local environment).⁸ Both (WS) and (SS) imply the existence of psychophysical correlations as the informal summary I provided implied. From the above formulations (WS) implies laws of the form $Gx \rightarrow Fx$ and (SS) implies laws of the form $Gx \square \rightarrow Fx$. (WS) is consistent with reading the correlation between the subvenient and the supervenient as merely accidental; and most philosophers would think that accidental covariance of properties is not a strong enough relation to ground a metaphysically robust form of dependency. Certainly, this is Kim's view. Hence, the only relation that appears as though it might ground dependency is (SS) which implies the existence of at least nomologically necessary connections between Fx and Gx . Let us consider this type of supervenience in more detail.

Kim in his 1989 paper, "The Myth of Non-reductive Physicalism" suggests that the existence of such laws should lead us to expect reductions. It is, though, quite interesting to follow through the steps of Kim's argument to see how his own analysis leads to a rather surprising result and provides an illustration of a general problem for non-reductive physicalism.

Kim assumes the Nagel model of reduction in which it is sufficient to reduce one theory to another if the reduced theory may be deduced from the reducing theory plus appropriate bridge principles. Kim's idea is that given (SS) we may use laws of the form cited above as our bridge laws in Nagel-style reductions, either singly to yield local reductions or jointly (and disjunctively) to

⁸ I think that Post is scared off (SS) because he takes it to imply individualism, that is narrow content. (An issue which I shall discuss directly in chapter 3.) This is obviously not the case, even if historically advocates of (SS) have also been believers in narrow content, as my extension of (SS) to cover Horgan's (1993) notion of regional supervenience shows.

provide us with a global reduction of some special science discipline to physics.⁹ However, I think it pays to look more closely at the sort of relation between the mental and the physical implied by the (SS) relation. One can best appreciate the sort of picture Kim has in mind with the following diagram:



The diagram represents the general structure of mental causation on Kim's theory. We take the horizontal arrow to indicate causation and the vertical arrows the dependency relation that is described by (SS). Only P-type events directly cause other P-type events, given the completeness of physics; and M-type events supervene on P-type events. It seems to be perfectly consistent with this picture that we imagine the M's to be *caused* by the P's. In other words, we may imagine the diagram represents the relations a form of dualistic epiphenomenalism, rather than a form of physicalism. Indeed, it seems that is exactly what it does represent, if we suppose that the mental property is not *identical* to a physical property and we follow Kim in identifying an event as an exemplification of a property at a particular time. In that case M and P represent different events and if we take the (SS) relation as a causal relation, we have a classic form of epiphenomenalism.

Kim, however, places an extra condition on a property being real which would seem to rule out construing the supervenience relation as a form of

⁹ This will allow Kim to overcome the problems multiple-realisation supposedly creates for reduction. In his writings Kim prefers local reductions, fearing disjunctive predicates will not be

dualistic epiphenomenalism. He claims that in order for any natural property to be construed realistically it should do some causal work.¹⁰ Since, by construction, one would think epiphenomenal mental properties or events do no causal work, they must be ruled out as real by this strengthening condition. But Kim's own analysis of what makes higher-level states causally efficacious is revealing. In the original paper, he maintains that one can reasonably claim that some mental event caused some physical event if the mental event in question *strongly supervenes* on the physical base which would feature in a complete *physical* explanation of the cause of the physical event. In terms of the diagram above, we could say then that M *caused* P' for example on the Kim model. Not because it does so directly: a direct diagonal causal arrow from M to P' is ruled out because if M causes P' directly that would violate the completeness of physics; or if it does not then it would at least mean P' was causally overdetermined which Kim thinks is unacceptable. Rather, because it stands in the correct (SS) relation to the physical cause, it itself may be considered a cause. Kim sometimes refers to this as supervenient causation. So the above diagram is an acceptable form of supervenient causation by Kim's lights. That is all well and good but we have yet to see any reason not to interpret the diagram as describing a form of dualistic epiphenomenalism. All that Kim's analysis has added is the counterintuitive idea that if one is a dualistic epiphenomenalist then one can claim Kim-style that mental events are causally efficacious because they strongly supervene on the physical.¹¹ Moreover, Kim still maintains that the relation between the mental

projectible and hence inapt for natural laws.

¹⁰ Kim calls this *Alexander's dictum*

¹¹ The argument can probably be run with a version of psychophysical parallelism as well. However, it might be objected since that theory of the mind-body relation severs any connection between the two domains it would only be strong enough to support (WS), not (SS). (Although it is not obvious that must be the case.)

properties and the physical properties is strong enough to support reduction. So, according to Kim although it would appear non-reductive physicalism is not a tenable position, given the various constraints he has placed on the mind-body relation, what we might call *reductive dualism* is perfectly acceptable.

I think this must show there is something deeply wrong at the heart of Kim's argument. The most obvious culprit is the Nagel-model of reduction. Simple deducibility of one theory from another plus "bridge laws" does not capture the intuitive idea of reduction. But that is not the most worrying problem for the advocate of supervenience. After all, most contemporary physicalists wish to resist the idea that the mental reduces to the physical. The disturbing thought is that the relation (SS), the strongest form of supervenience, is not by itself strong enough to rule out forms of dualism. Even that is only half the problem. The very fact that the kind of dualism supervenience plus the completeness of physics permits is a version of epiphenomenalism seems by analogy to lead to the conclusion that the mental is causally inert *no matter how* the supervenience relation is interpreted. If it does not, in other words physicalists can explain how the mental can conform to the pattern of (SS) and still be causally efficacious, then the possibility that dualists may use the same form of explanation to account for the efficacy of substantially different mental properties and events is opened up. Either way, one, perhaps the main, motivation of physicalism is undermined: its claim to remove the problems of mental causation that have plagued dualists since Descartes.¹² This is why I think the question of reduction matters little one way or the other. The non-reductive physicalists needs an account of the relation between mental properties, states or events and physical properties, states or

¹² Tim Crane (1995) makes the same point.

events which rules out dualism and rules in mental causation. The standard account seems to do neither.

Kim and many other writers on this subject are aware of these problems. For example, Kim (1993) has written that:

Mind-body supervenience, therefore, does not state a solution to the mind-body problem; rather it states the problem itself. (p.168)

And

But the Thesis itself [supervenience] says nothing about the *nature* of the dependence involved: it tells us neither what kind of dependence it is, nor how the dependence grounds or explains the property covariation.... When we reflect on mind-body supervenience and compare it with the traditional options, we are struck by its failure to address the explanatory task. (p.166-7)¹³

In fact, Kim has invented the term, *causal exclusion arguments*, for the general sort of argument we have considered which seems to lead to the conclusion that the mental is epiphenomenal. This is a problem which he takes very seriously and has dominated his recent writings.

Clearly then the notion of supervenience as it stands is not adequate to ground a form of non-reductive materialism. One reason for this is that the technical notion of supervenience is non-symmetric. Strong supervenience implies

¹³ See also Horgan (1993) and Heil (1998)

that supervenient properties covary (admittedly necessarily covary) with subvenient properties, whereas the intuitive idea behind non-reductive physicalism is that the physical *determines* the mental. Determination should, intuitively, be an asymmetric relation. Moreover, when one reflects upon the conventional position it is a strange mix of views. It involves a commitment on the one hand to monism, since everything is fundamentally physical, and on the other hand property dualism (or pluralism), since the world contains irreducibly non-physical properties. What non-reductive physicalists need to combine these monistic and pluralistic parts of their view is an explanation of why the mental supervenes on the physical which is physicalistically kosher; an explanation that converts the standard non-symmetric relation of supervenience into an asymmetric relation of dependence and addresses the apparent problems of causal exclusion reasoning. What is needed, in Terence Horgan's words, is superdupervenience.

2.2 Covariance to determination

How is the move from supervenience to superdupervenience to be effected? Different philosophers have different ideas about the best way to go. These, I believe, can be grouped into three types of solution. First there is the idea that the relation between sub and supervenient properties should be modelled on that of the relation of parts to wholes. Second that the relation between determinants and determinables provides the paradigm for an acceptable version of supervenience. Third it has been claimed that by analysing mental states in terms of their causal powers, an acceptable account of the supervenience relation can be constructed. To assess the plausibility of each of these claims two kinds of question must be

addressed: what I shall call *internal* and *external* questions. The internal questions concern whether or not the relation between the sub and supervenient properties is acceptable by physicalist standards; and whether construing the relation in that way removes the apparent problems created by the causal exclusion argument. The *external* question is whether the relation so described is a plausible description of the relation between physical and non-physical (in particular mental) states.

Take for example the idea of property identity as an explanation of the supervenience relation. It clearly satisfies the internal questions. If mental properties and thus events and states are identical to physical properties it is unmysterious how mental events cause physical events. They *are* physical events. However, modern scruples regarding multiple-realisation mean this suggestion fails for most philosophers to answer the external question. It is implausible that mental states are identical to physical states. We shall consider the same questions for the three versions of superdupervenience.

2.3 A model for superdupervenience I: parts and wholes

First consider the idea that the supervenient properties are related as parts are to wholes, what Kim calls mereological supervenience (MS). Perhaps something like (MS) plays a role in many people's intuitive understanding of the supervenience relation. The analogy between parts and wholes is obviously a tempting one, one which fits naturally into the hierarchical picture of the universe which forms the background to much contemporary physicalist thinking. Moreover, intuitively one would not think of parts and wholes in causal competition with one another. That is to say, it would hardly be credible to claim, for example, that because the bricks

and mortar caused the ball to change direction that ruled out the wall, composed of those bricks and mortar, from causing the change in the ball's flight. Worries here about epiphenomenalism would simply be misguided.¹⁴

What an advocate of (MS) requires is some general account of the relation between parts and wholes that could serve to flesh out these intuitions and provide an account of supervenience. The most obvious method (for a philosopher) would be to employ some form of mereological logic as, for example, developed by Nelson Goodman. Subvening properties would be characterised as proper parts of supervening wholes. However, this logical characterisation of (MS) would be at once too restrictive and too permissive. Too restrictive because there are obvious examples where classical mereological logics fail to relate parts to wholes in, intuitively speaking, an illuminating way. For example, biological entities are such because they have a certain evolutionary history; an organism summed from the same physical parts but with a different selection history would be a different type of species. But that is not a distinction that can be made using a classical mereological logic. Similar points could be made regarding artefacts and social institutions. The properties of such entities that are special, that help identify them as what they are, cannot be captured by the simple idea of mereological fusion. Wholes like artefacts, social institutions and biological entities have properties that we would not attribute to a mere mereological sum of their parts. On the other hand, mereology is also too permissive. If the idea of mereological supervenience is merely the logically summing of parts, then (MS) will be trivial and ubiquitous.

¹⁴ There are some general conceptual problems regarding the relation between parts and wholes which need to be cleared up, most importantly the problem of vagueness, before any (MS) relation could get going. One can think of the classical sorites paradox as a paradox of mereological composition. That is to say, wholes are relatively and to a vague degree insensitive to changes in their parts (or at least some wholes are). Trenton Merricks (1998) objection to microphysicalism

There will be thousands and thousands of entities composed of arbitrary parts. There could be no change in any mereological sum without a change in its elements and conversely no change in the elements from which entities are summed without change in some (many in fact) arbitrarily composed mereological sums. So the intuitive asymmetry of a satisfactory supervenience relation would be lost.

Indeed, although it might be tempting to start there, I do not know of any philosopher who has actually offered an account of non-reductive materialism just using a mereological logic.¹⁵ What is needed to give (MS) some bite is an explanation of *how* it is that the *properties* of the parts combine and give rise to the *properties* of wholes.

Gabriel Segal and Elliott Sober attempt something like this by defining (MS) as follows:

(MS) M mereologically supervenes on P iff it is *nomologically necessary* that if any object, x, has P at any time, t, then x has M at t.¹⁶

(1991, p. 10)

The relation that Segal and Sober are after is some kind of lawlike connection between M and P. Insisting on a nomological connection between parts and wholes should discriminate between the significant and interesting supervenience relations and the arbitrary conjunctions of Goodman-style logics. Moreover, such lawlike connections might help us understand the relation between the properties

turns on this fact. Nevertheless, I would prefer not to make my argument against physicalism on a par with an argument that said heaps of sand were *not* constituted by grains of sand.

¹⁵ It does form part of Hellman and Thompson's (1972) definition.

¹⁶ I have adjusted the notation to fit with my diagram. The italics are mine.

of, say, artefacts and the properties of their parts that the mere summing of parts cannot. However, there are some familiar problems here. First of all, on the conceptual level, it is not clear why such connections are physicalistically respectable. After all, since such a law involves citing a physical *and* a mental property, it cannot be a physical law. Moreover, one could hardly claim that such a law supervenes on a physical law because again we would need an explanation of what was meant by supervene. If we simply invoked another 'nomologically necessary' connection between this law and the supervening psychophysical law then we would have the beginnings of an infinite regress. That is to say, since the psychophysical law (call it P1) itself is not physical it must supervene on something which is. Call the supervenience relation which relates P1 to a physical law, P2. P2 is not physical either so it must supervene on something physical. Call this new supervenience relation P3. P3 is not physical either so... and so on. Unless we are to understand these psychophysical laws in a special way, which can avoid the need for further explanation that I have adumbrated above, Segal and Sober have failed to explain supervenience. In fact, they offer no such account. So it remains unclear to me why what Segal and Sober have presented is not just a version of dualistic epiphenomenalism or emergentism, as described earlier, in which non-physical laws relate physical to non-physical states.¹⁷

In addition to their failure to explain the supervenience relation adequately, Segal and Sober provide an equally unsatisfactory solution to the problem of mental causation. What they require for a mental state to be efficacious is that

¹⁷ It might be thought that the condition that the properties be instantiated in the same individual rules out dualism since substance dualism implies the existence of two individuals. However, various emergentist views would not be ruled by this condition. Classical British emergentists thought there were precisely these kind of physical-non-physical laws which were basic and irreducible.

there exists a law¹⁸ connecting the supervening properties (M and M' in the diagram above). But it is unclear what such a condition amounts to without some analysis of what makes a lawlike connection, a lawlike connection. For example, if all that is meant by lawlike is that there exists some regularity pertaining between M and M', Segal's and Sober's condition can easily be fulfilled by the sort of epiphenomenalism I have described. For example, it may be that there are regular connections between M and M' because P *causes* M, P' *causes* M' and there are regular connections between P and P'. The point holds generally as long as we consider the lawlike relation between the M's to derive from the lawlike connections between the P's; and given that the M's are supposed to supervene on the P's this seems inevitable. However, because we have no explanation of the mental to physical connections such that they can rule out certain kinds of non-physicalist positions, it looks like Segal's and Sober's account if it worked would also allow epiphenomenal dualists to claim that mental states *really* are causally efficacious. This must show it can't work.

To put the same point another way, any law connecting mental states looks like it will supervene on physical laws connecting the physical bases on which the mental states supervene. If the M-law does supervene on the P-law or laws, then we need an account of that supervenience relation, one which is physicalistically respectable. As I argued above, Segal and Sober have failed to provide such a general account. On the other hand if the M-law is not derivative then there would seem to be a non-physical fact (a true law) which does not supervene on the physical; and it is difficult to see how that could be consistent with physicalism. In short, just as Segal's and Sober's account of supervenience is inadequate

¹⁸ The actual condition is a little stricter than this to rule out grue-type cases. The laws must be "useful", meaning here as far as I can see just that they are the sort of laws scientist might actually

because it fails to explain how the supervenient connection fits into a physicalist ontology, so their account of mental causation is inadequate because it fails to explain how invoking lawlike connections between non-physical states can be compatible with physicalism without being supplemented by some further unexplained supervenience relation.

A possible means of giving (MS) more empirical bite than Segal's and Sober's account allows would be to consider some concrete cases of micro-macro connections. One could then argue that the appropriate, explanatorily robust connections between sub and supervening properties¹⁹ were more of the same sort. Terence Horgan makes essential this suggestion:

For at least some kinds of property we have a fairly good idea about what would count as a materialistically acceptable explanation of why such a property is supervenient on a given configuration of physical properties... We understand well enough the essential features of liquidity... Thus explaining why liquidity supervenes on certain microphysical properties is essentially a matter of explaining why any quantity of stuff with these microphysical properties will exhibit these macro-features. (1993, p.579)

But even if we grant that there exists an explanation in the above case, it provides no *general* account of (MS). At best, it provides a general aspiration: that there should be explanatorily robust, physicalistically kosher explanations between the sub and the supervenient states. What is needed in any particular

use.

¹⁹ Jeffrey Poland's (1994, p.209) fourth thesis of physicalism is essentially a version of this.

case, if one is to plausibly claim that any (MS) relation holds, are *actual* explanations of the relations between sub and supervenient properties. It is clear, though, that in many cases we do not have anything like that as the example of quantum chemistry I discussed in the previous chapter indicated.

The prospects look gloomy for any piecemeal account of supervenience along these lines, as Horgan himself admits (1993, 581-2). Whether or not the mental mereologically supervenes on the physical in this way is at best idle speculation about a we-know-not-what relation.²⁰

One final point should be made against the whole idea of (MS) as a *generally* adequate account of supervenience. The relation between the mind and the body is for modern day physicalists a relation between brain states (plus maybe something else) and mental states. It is difficult to see in what way that relation can be said to model the relation between parts and wholes. It is frankly incredible to say that bits of the brain add up to make up wholes which are thoughts. Bits of the brain only seem to add up to more brain. Moreover, it is difficult to see how the properties thought to be definitive of the mental, its normative and intentional aspects can be thought to relate to the physical as wholes do to parts.²¹

²⁰ Nor would it help I think to say that the connection between sub and supervening properties is a physical law. First because it is difficult to see how any non-physical property could be the result simply of the combination of physical parts according to physical law. That would surely at best give rise to just a physical whole. But also second and more seriously, it is not clear what invoking the idea of a physical law would amount to here without some better understanding of what does and does not count as physics; and again the point should be made that contemporary science raises serious concerns about the general viability of such claims. See the discussion of quantum mechanics in chapter 4 for a detailed discussion of the problems.

²¹ Kim (1998,1999) has emphasised this point. Sober has informed me in correspondence that his characterisation of (MS) is not an attempt to model supervenience on the part-whole relation. Rather he is suggesting that certain properties supervene on physical wholes, the properties of which are in turn dependent on the properties of their parts. For Sober then such supervenience relations that exist between say the brain and the mind remain unexplained.

In summary then, mereological supervenience fails to provide satisfactory answers to either the internal or external questions. Understood as a logical thesis, it fails to explain the relation between parts and wholes in an illuminating way. Stated vaguely as a nomological connection it is unclear why it rules out various dualist positions or how it can avoid a possible infinite regress. Put forward as a suggestion that there needs to be detailed explanation between the micro and macro parts which demonstrates that the relation is physicalistically kosher, it suffers from the obvious defect that we have no such explanations in the cases that we are interested. Furthermore, there has been no satisfactory response to how this account of supervenience avoids the causal exclusion problem. All of the internal problems from which supervenience suffers remain. Moreover, it is clear after a moment's reflection that the relation between parts and wholes does not describe many of the supervenience relations in which we are interested, in particular the mind-body relation.

2.4 A model for superdupervenience II: determinates and determinables

Let us turn then to other views. Stephen Yablo's is undoubtedly the most interesting. He believes that the relation between sub and supervenient properties and hence the relation between mind and body should be thought of as a species of the determinate-determinable relation.²² That is to say, mental properties are related to physical properties as the property of being red is related to the property of being scarlet. *Prima facie*, this seems most implausible since the relation between a determinate and its determinable is a conceptual one, whereas that

clearly does not seem to be the case for the relation between the mental and the physical. To soften us up a bit, Yablo appeals to the well-known examples of *a posteriori* identity claims – water = H₂O, etc. He then reasons as follows: if there can be identities which are only known *a posteriori*, then there can be relations of determinate to determinable which are equally *a posteriori*. Just as we may be deceived into thinking that water and H₂O are distinct, so might we be deceived that the mental is not a determinable of certain physical determinates. I think the initial intuition carries more weight than Yablo allows here but it is worth discussing his ideas in greater detail; in particular, how it is he thinks he can avoid epiphenomenalism.

First, let us begin Yablo's definition of the relation between determinates and determinables:

(Δ) P determines Q iff: for a thing to be P is for it to be Q, not *simpliciter*, but in a specific way. (1993, p.252)

The specific way being that described by the determinate-determinable relation: to be scarlet is to be red in a specific way. As with strong supervenience, if P is instantiated, then Q must be also but Q may be instantiated without P; that is to say, such an interpretation of the mind-body relation looks open to the possibility of multiple-realisation.²³ This leads naturally to the following analogous relation for events:

²² Segal and Sober's (1991) complete account arguably employs a version of this notion as well. This is how they make air being present efficacious for the lighting of match even though only the oxygen in the air is necessary.

²³ See below though.

(δ) p [event] determines q [event] iff: for p to occur (in a possible world) is for q to occur (there), not *simpliciter*, but in a certain way. (p.260).

But since events are particulars some sense needs to be given to the idea that they may occur in a certain way; in other words, what it is to be a particular event needs to be explained. Yablo does so by introducing the notion of the essence of an event. Technical complications aside, an essence is the set of properties necessary for the existence of the event, excluding trivial analytic properties and those which are consequences of the basic and (hopefully) logically independent set of essential properties.²⁴ This is a variation on the Kim model of events. An event is understood as an instantiation of a cluster of properties, its essence. On Yablo's account physical events 'subsume' mental events; that is the essence of the mental event is determined by the essence (or part of the essence) of the physical event. So far this is just a more explicit version of the standard supervenience account. Mental events and physical events are non-identical, they have different essences in Yablo's terminology, but physical events determine mental events. The original part of Yablo's claim is just the thought that the determination relation is to be explained as a species of the determinate-determinable relation. In other words, the essence of any particular mental event is the determinable which is subsumed by a physical, determinate essence. Epiphenomenalism is then side-stepped since "determinates do not contend with their determinables for causal influence"(p.259). For example:

²⁴ For more technical detail on the properties that make up an essence see Yablo, p.262, n.37

[Archimedes] shouting “Eureka!” was causally sufficient for his cat’s startled flight, nobody would think that this disqualified his (simply) shouting from being causally relevant as well. (p.272)

In fact, Yablo goes further than simply saying that determinates do not compete with their determinables for causal influence. There should be, he claims, a “fit”²⁵ between causes and their effects. “[T]hey should incorporate a good deal of the causally relevant material and not too much of the causally irrelevant material...the cause was the thing that made the difference between the effects occurring and it not.” So, in some situations it is better to say the determinable caused the event and let the determinate with all its extraneous detail drop out of the picture. Understanding the determinable event then as a cause of another determinable leads to this kind of picture:

Determinable (Socrates drinking the hemlock) → Determinable (Socrates death)

↑

Determinate (Socrates guzzling the hemlock)²⁶

Since Socrates *drinking* the hemlock is sufficient to cause his death, being told that Socrates *guzzled* the hemlock only adds unwanted and unneeded detail.

To summarise: Yablo’s account makes use of two principles to overcome what I have called the internal problems of supervenience. First, supervenience is a version of the determinate-determinable. This clearly renders the supervening, in

²⁵ Yablo employs the term ‘proportionality’.

²⁶ In case this is not obvious: the vertical arrow is the supervenience relation here given by the determinate-determinable relation; and the horizontal arrow represents the causal connection.

Horgan's words, physicalistically kosher. The determinate-determinable relation is not a mysterious or emergent kind of property relation. Second, Yablo's account of causation, in which the cause of one event brings about the essence of its effect, combined with his analysis of supervenience overcomes the causal exclusion problem. It follows on this account that sometimes the determinable (that is the supervening event) plays the role of cause.

Yablo's account is certainly interesting and I think it succeeds in avoiding epiphenomenalism for higher-order properties and thus events. However, he fails to convince me of his central thesis that the relation between the mental and the physical is a version of the determinate-determinable relation. I just find that implausible. But what say you (and Yablo too) of the analogy between H₂O and water? In the following I will try to articulate my reasons from my dissatisfaction with Yablo's account.

First one should note that Yablo's solution to the mind-body problem seems impossible to combine with certain theses concerning the mental that have wide currency at present. For example, if you agree with Searle that there is something irreducibly subjective about the mental, then you are unlikely to be persuaded that the determinate-determinable relation is the right way to describe the connection between the mental and the physical. The relation between third-person and first-person perspectives does not seem to be a species of the determinate-determinable relation. If you hold that the mental involves irreducibly rational-normative qualities, then it seems unlikely that such an is-ought gap can be bridged in the way Yablo suggests. Similarly, if you hold that mental properties are historically-based, that is to say the content of mental states in some way depends on the individual's causal history, then similar problems arise. The

determinate-determinable relation seems inappropriate to bridge the gap between the physically ahistorical and the non-physical, historical properties.²⁷ Perhaps, though, Yablo rejects such subjectivist or normativist or historicist views of the mental. If so, then the onus is on him to tell us how the mental should be characterised so as to fit his model.

Even if we are willing to be charitable here, the analogy between determinate- determinable relation and supervenience is not a perfect one, as Douglas Ehring (1996 p473-474)²⁸ has pointed out. Yablo, like most philosophers, accepts the idea that mental states are multiply-realizable; and, as I explained in my sketch of his views he takes it that the determinate-determinable relation captures this idea. Indeed it is the fact that the determinate-determinable relation corresponds so closely to many of the intuitive ideas behind non-reductive physicalism that leads Yablo to recommend it as an interpretation of supervenience. However, Ehring demonstrates that if we press the analogy with determinates and determinables we discover that in reality it is not consistent with standard ideas about multiple-realizability.

Consider again a classic example of the determinate-determinable relation. Red and blue are both determinants of the determinable colour. Ehring points out that as determinates, they are different with respect to their determinable: they are different colours. This is a general feature of the determinate-determinable relation. Now consider the usual multiple-realizability story. Two individuals have the same belief but with different physical realisers. On Yablo's theory the realisation relation is the determinate-determinable relation. As we have seen

²⁷ A similar point may be made with regard to externalism. That said, Yablo (1998) has an excellent paper which argues that wide-content is efficacious. However, this argument depends on Yablo's analysis of causation in terms of essences. He fails to make it clear how we can think of wide-content states being determinables of physical determinates.

determinates that are different are different with respect to their determinable. Hence, since the two individuals have different realisers, they must have different beliefs, which contradicts the initial assumption of multiple-realizability.

Yablo could of course reject multiple-realizability. But we must remember that the attraction of Yablo's account is supposed to be how closely it conforms to the intuitive ideas behind supervenience and all that comes with it, including multiple-realisation. Once Yablo's account begins to move away from those intuitions and perhaps more importantly the arguments which support them, his case for the determinate-determinable relation correctly describing the mind-body relation is weakened.

Perhaps these considerations are not sufficient to abandon Yablo's programme. Yet even if Yablo denies multiple-realizability, the normative, historical and subjective elements of mental life, I believe his view still encounters problems when we turn to the one property which is undeniably a characteristic of the mental – intentionality.

Let us consider more closely the determinate-determinable relation. A feature of the relation is that relates a more specific with a more abstract property. In general, we proceed up a hierarchy of determinate-determinable relations. For example, we proceed from scarlet to red to colour to, perhaps, ultimately, property. (We might say the determinable property has all other properties as determinates.) What we expect from properties that stand in the determinate-determinable relation is that they should exhibit this general structure. A determinable is more abstract than its determinate. When we turn to consider intentional states they do not seem to be in that sense more abstract than the

²⁸ He has others objections to Yablo's account which I find less convincing.

physical states which putatively realise them. Given a physical realiser, call it X, couldn't we imagine a belief state which is equally specific: namely the belief that X? Of course, this is not to suggest that the actual realiser of the belief that X is X, just that whatever the realising state is, it will be no more specific than X. If that is possible then mental states do not exhibit the right kind of relation to physical states to be determinables of physical determinates.

Perhaps this is impossible. Perhaps one cannot have a mental state which has content as specific as the state which realises it. (In that case, assuming physicalism is true, it would be impossible ever actually to discover the physical state which gave rise to a particular mental state. Such a state being too specific for our minds to cope with.) Even, if the argument falls short of a refutation, it does highlight how counterintuitive Yablo's claims is. When we consider the rich detail of intentional states, it is difficult to believe that they are abstract enough to figure in the kind of determinate-determinable relation that Yablo claims they do.

The combined weight of the foregoing considerations make the case against Yablo's theory a powerful one, particularly since he provides no detailed attempt to articulate how the determinate-determinable relation would actually apply to the relation between the mind and the body. Without an account of the mental which enables us to see how the determinate-determinable relation could be relevant in describing the connection between mind and body, and given the above reasons to think that there could not be one, I think we should reject Yablo's theory. At the moment, all we have is an interesting and *partial* analogy between the formal notion of supervenience and the determinate-determinable relation; one which admittedly *in the abstract* solves the problem of the causal exclusion argument. However, there are obviously other necessitating

relationships between properties that are *not* determinate-determinable relations which might do the same for us. One possibility that plagues all non-reductive accounts is causation but there are others. For example, red is a determinate of the determinable colour: nothing can be red without also being coloured. It is also true that nothing can be red without also being extended in space. Redness and extension are quite different properties, why shouldn't this provide a better model for the mental-physical connection?²⁹ If Yablo is to persuade us his proposal is right, he needs to provide some positive reasons for favouring his account over others like the red-extension example. He also needs to be able to explain away the apparent disanalogies between mental-physical relations and determinate-determinable relations that have been highlighted in the foregoing. He needs to provide some reason for believing that mental-physical connections are a species of determinate-determinable connection. So far we have nothing like that. Thus I believe there is no reason to accept Yablo has given an adequate account of superdupervenience.

Cynthia and Graham Macdonald (1995) offer a solution to the problem of mental causation very similar in spirit to Yablo's but which takes a different and important line on causation. Yablo and Kim take mental events to be different from physical events and this raises serious concerns about dualism, which we have explored at length above. However, one need not accept this; and indeed the Macdonalds do not. They advocate a token-token event identity thesis. That is to say, each concrete instance of a mental event is identical to a physical event but not type identical.

²⁹ Would this be compatible with physicalism? I have no idea because as with Yablo's account I can't see how to apply it to the mind-body relation.

The Macdonalds also draw an important distinction between causation and causal explanation. Causation, they claim, is an extensional relation between *events*, explanation an intensional relation between *properties*. The Macdonalds are therefore in a position to distinguish *three* varieties of epiphenomenalism, all of which have to be avoided if non-reductive monism is to be a viable position.

1. That mental *events* are causally inefficacious
2. That mental *properties* are causally irrelevant
3. That mental *properties* are *explanatorily* irrelevant

Presumably any account that deals with (1) and (2), should have little difficulty in tackling (3).

The Macdonalds' account, with its emphasis on the extensionality of the causal relation and advocacy of token-token identity has some very obvious similarities to Donald Davidson's Anomalous Monism. One important area where they diverge from Davidson, however, is in their account of events. Here they follow Kim and most other modern advocates of the supervenience relation in defining an event as an exemplification of a property at a particular time.

Let us consider now how these views on events, even identity and causation and explanation fit together. A token-token identity thesis for events takes us a long way to answering epiphenomenalism 1. If some mental event causes some piece of behaviour, then it must be identical to some physical event – if the physical event is causally efficacious then the mental event must be too. On this account then we don't need the determinant-determinable relation to make

sense of the connection between mental and physical *events*. Mental events just are (that is, token-identical to) physical events.

So far so good. It seems relevant to ask then what makes any event, a cause. According to the Macdonalds (again closely following Davidson) an event is a cause if it is an instantiation of a physical property that falls under a strict (i.e. exceptionless) law. Troubles are clearly looming. If it is *in virtue* of the fact that the event falls under a *physical* law that the said event is efficacious, then it seems natural to say that only the physical event or the physical part of the event is involved in causation.³⁰ One might put this by saying that the mental event *qua* mental event is epiphenomenal, the cause is the physical event *qua* physical event. The Macdonalds' response is to claim that: "two properties, one mental and one physical, can be jointly instanced in a *single* instance (i.e. in an individual event). The mental property will then have an instance which *is* (i.e. identical with) a nomological [i.e. physical] property." That is to say, they invoke their token-identity thesis. Like Davidson they claim that to talk of events causing one another *in virtue* of certain properties is to confuse explanation with causation. To repeat, on this account if any token event, any particular occurrence of a physical property is causally efficacious, which again it surely must be, then, since the particular instancing of the mental property is *identical* to that physical property instance, it must too be causally efficacious. And since the instancing of the mental property is what we would call the event, epiphenomenalism 1 is avoided.

At least so the Macdonalds say but some pretty strange ontological claims are made in this account. What can it mean here to say that *property-instances* are identical? How can token instances of *properties* be identical, without being *type*

³⁰ This is, of course, a standard of criticism of Davidson. See below, chapter 7, for why I don't think it applies to his version of Anomalous Monism.

identical? Well, of course, if we construe each property instance as an instantiation of a universal, then they can't really. Mental properties are different from physical properties so instantiations of mental properties must also be different from instantiations of physical properties. Intuitively, either properties or property instances are type-identical or they are not identical at all.

It is here that we see the importance of the Macdonalds' break with Davidson. Like Davidson the Macdonalds wish to claim that causation is only a relation between events and since, *ex hypothesi*, the mental event and physical event are the same then no problem regarding epiphenomenalism should arise. However, Davidson can pull off this trick because events *qua* events are neither mental nor physical, only the predicates used to pick the events out have these characteristics.³¹ The Macdonalds, though, do not adopt a Davidsonian theory of events, nor do they follow his nominalism regarding properties – events remember are objects instancing a *property* at a certain time. But since mental properties and physical properties are obviously *different*, I repeat, it surely follows that instances of mental properties at a time are different events from instances of physical properties at a time. And if they are separate mental and physical events, then epiphenomenalism re-emerges as a problem for the Macdonalds. Given that it is only in virtue of falling under a strict physical law that any event is efficacious at all, if mental events are different from physical events since they don't fall under strict laws, they must be epiphenomenal.

The alternative to this conclusion would be to insist that *all* the properties instantiated by an object at a time *t* are part of the same event. But that is going to lead to some frankly absurd event identity commitments. For example, I, as an

³¹ See ch. 7 for a fuller discussion.

object, currently instantiate the properties of typing and of feeling hungry. Surely no-one would claim that the typing event and the hunger-feeling event were one and the same? Yet that appears to be exactly what the Macdonalds would have to say in order to avoid epiphenomenalism 1. So it seems that *either* the Macdonalds cannot claim that the mental event is causally efficacious because they have no grounds for saying that the mental event and the physical event are identical *or* they must adopt a highly counterintuitive and profoundly implausible account of event identity. If the Macdonalds are to avoid both of these possibilities they must be able to provide a principled account of when we would say that two *different properties* are part of the *same event*; and convince us that the relation of the mental to the physical is such a case

The Macdonalds do indeed offer an example of when we might *reasonably* say two different properties are coinstanced in the same event. What is it? None other than Yablo's example of determinates and determinables. If a particular event is the instancing of red at a particular time, then the very same event is also the instancing of colour at the same time. (p.65)³² But we have just seen that the determinate-determinable relation does not satisfactorily characterise the relation between the mental and the physical. In fact, the Macdonalds concur. The example is simply used to demonstrate that there is no absurdity in claiming that one event may co-instance different properties.³³ I gladly acknowledge there is no absurdity here. In the case of determinates and determinables it is very clear why this is possible: there is a conceptually necessary connection between a determinate and its determinable. (This is what the Macdonalds say too.) The

³² Of course, on Yablo's account we should still construe the mental and physical events as different even though they stand in the determinate-determinable relation.

relation between physical properties and mental properties is not *prima facie* a conceptual one. How then can they be accommodated into this picture? Because, so the Macdonalds claim, the mental *supervenes* on the physical; and that is another way, a *metaphysical* way, in which two different properties can be co-instantiated in the same event (p.66). Now we are right back where we started. We want to know *how* it is that supervenience is supposed to guarantee this. In particular, why the coinstantiating and covariation of properties renders them causally relevant.

It may seem that a solution to epiphenomenalism 2 could help us out of our difficulties with epiphenomenalism 1. If the Macdonalds could provide clear reasons for explaining the causal relevance of mental properties, then that might go some way to alleviating the worries I have raised about their token-identity thesis. That is to say, in explaining the importance of the mental *property qua* causally efficacious event, that is its relation to the physical property, we might be able to construct some sort of theory which can make sense of the Macdonalds' claim that the token physical property instance and the token mental property instance are one and the same.

Unfortunately, the Macdonalds' solution to the question of the causal relevance of the mental event is disappointingly familiar. Like Segal and Sober, the Macdonalds claim a property is causally relevant if it is part of an efficacious event and, in virtue of the fact that its property type falls under some kind of pattern (in the case of mental states this is the pattern of rationality). Again things seem mighty queer. The fact that the individual instantiation of the event falls under some pattern or other seems to do nothing to remove the threat that all the

³³ In actual fact, the Macdonalds offer up other examples – functional properties. It does not seem though they endorse this understanding of the mental either. See the discussion of Kim below,

causal work has been done by the physical properties. What makes an event efficacious in the first place has to do with, and only to do with, the physical properties of the event. It seems impossible to avoid the conclusion that the mental properties are causally superfluous unless you follow Davidson down the path of nominalism and deny that *any* properties (or predicates) are causally significant beyond picking out events which are causally related.

I just cannot see how to tie together the Macdonalds' token-identity claims with their view on events and the causal relevance of properties. Because there is no worked out story of how the mental and physical relate, just an open-ended appeal to supervenience, there seems no reason to believe that the mental aspects of any event are anything other than superfluous to causal requirements.

One way of avoiding some of these strange claims about event identity and retaining some of the benefits of the token-identity view is offered by John Heil (1992). Instead of thinking of the event as a property instantiation, he takes it to be a trope, an abstract particular. This has some obvious advantages over the Macdonalds' account. Because tropes are already particulars there does not seem the same tension between accepting event identity and rejecting property identity. Nevertheless, there are some similar problems on the horizon once we consider the question of how to construe tropes.

Tropes normally are thought of as providing a sparse theory of properties (see Campbell (1990), ch.1). In other words, there exists a set of tropes which stand in a relation of exact similarity (i.e. qualitatively identical but numerically distinct) corresponding to every property required by a complete science (i.e. physics, for a physicalist). The properties picked out by these elite classes feature

ch.3, for the drawbacks of a functionalist approach to these matters.

in genuine and complete laws of nature. Hence, a trope must be individuated on this theory by its physical aspects. The sense in which these physical tropes can be said to be identical to mental property instances is somewhat *jejune*. Mental property instances do not directly correspond to any trope. The class of mental properties must be constructed out of other sparse tropes. But the mental property so constructed is *nothing more* than a construction: a classification of tropes that do not stand in a relation of exact similarity to one another. So on a sparse theory of tropes, there are only mental properties as a result of a construction from physical tropes – it is difficult to see how such constructed properties can be said to be genuinely causally relevant.³⁴

On the other hand, if we try to construe tropes as abundant, i.e. there is at least one class of exactly similar tropes for every predicate, then we run into even more serious problems. Why should we say the mental trope is identical to the physical trope, why are they not two different tropes? What are the identity conditions for tropes? It can't be spatial location, since different tropes are supposed to be able to instance different properties while being located in the same place. For example, a red box is the coinstantiation of a red trope and a cube trope (and possibly some others). All of these tropes must be in the same place: that is where the red box is. It can't be done by picking out the same cause and effects, since the causes and effects would also have to be tropes and an obvious circularity would threaten. That is to say, the individuation of tropes would depend upon some prior individuation of other tropes. So it looks like we have the

³⁴ The Macdonalds might be interpreted in an analogous manner. All events are strictly speaking physical events but they may fall under different patterns that allow them to fit into non-physical classifications; mental properties such as they are, are merely reclassifications of physical states. See the discussion of Kim's views below to see how this might form part of a functionalist view.

same problem that Macdonalds have: we have no way to make sense of the putative token-identity claim.

In truth, neither Heil nor the Macdonalds is able to provide a satisfactory characterisation of how it is the mental and physical events can be one and the same, while making the mental part of the event causally relevant in any but the most Pickwickian of senses. The benefits of some kind of token-token identity theory turn out to be a mirage.³⁵

2.5 A model for superdupervenience III: causal powers and the return of functionalism

Jessica Wilson (1999) claims that all we need to add to supervenience to render it superduper is the constraint that supervening and the subvening properties (or causal powers) be related internally to one another. Her paper offers one suggestion as to what that internal relation might be. Let us first consider the general claim that all that is needed to render the relation between the supervening and the subvening physicalistically kosher is that relation be internal, before going on to consider Wilson's own particular proposal.

Wilson follows Armstrong in defining an internal relation as one which "is dictated solely by the nature of the relata, and is such that given certain entities with certain natures, the relation must hold between the entities." Since it is in some sense metaphysically necessary that internally related properties supervene on one another there is no threat that the relations between the subvening and the supervening will not be physicalistically respectable.

³⁵ Davidson is a different matter, see chapter 7.

This is all right again as far it goes. But it seems to me doomed to failure as a general account of the supervenience relation since it is highly controversial exactly what is and what is not an internal relation. There are, of course, some obvious examples of internally related properties, self-identity for one. And there are probably some non-controversial examples of external relations, spatial and temporal distance for example. (Leibniz, would of course, demur.) But for the one relation which really matters – the causal relation – it is a point of contention amongst philosophers whether or not to characterise it as an internal relation. Indeed, sometimes disputes between Humean^s and non-Humean^s regarding causation are characterised as concerning whether causation is an internal or an external relation. Now if we can plausibly regard causation as an internal relation, then we clearly have a physicalistically unacceptable relation between the supervening and the subvening properties as this is exactly the relation that epiphenomenal dualists have in mind as we have seen. Moreover, as we have also seen, if all that is required to allay fears that the mental may be epiphenomenal is that the sub and supervening properties be internally related, then Cartesian interactionism can be made a philosophically respectable position so long as we construe the causal relation between the sub and supervening properties as an internal one.

Is there then a respectable construal of causation as an internal property? Ironically, one of the philosophers that Wilson quotes with approval – Sydney Shoemaker – provides an account of causation which is exactly that. He argues that no property would be what it is if it lacked any of its causal/conditional powers.

[A]ll of the causal powers possessed by a property ... are essential to it. This has a very strong consequence, namely that causal necessity is a species of logical necessity. (Shoemaker, 1984, p.222)

Hence, all its conditional powers stand in an internal relation to the property. So if one such conditional power is to produce non-physical states, then the relation is at once supervenient, internal and causal; and thus not physicalistically kosher. How much sense any philosopher thinks he can or cannot make of the idea that causation is an internal relation or talk of causal powers is not for the moment important. What is important to note here is that a simple quasi-logical/metaphysical distinction like that between internal and external relations will not help when trying to define superdupervenience. It does no good, for example, to say that the appropriate relation is internal but we should not consider causation an internal relation because the question immediately arises: why should we exclude causation, i.e. what *kind* of internal relation *is* supposed to be acceptable? To repeat, what is needed is a concrete and detailed explanation of the supervenience relation and why anyone should believe this accurately characterises the relation between super and subvening properties. An appeal to internal relations is a long way from this.

As well as offering a general diagnosis of what is wrong with supervenience, it fails to characterise the relation as internal, Wilson suggests a particular remedy. An adequate supervenience relation must obey the following principle:

Condition on Causal Powers (CCP). Each individual power associated with a supervenient property is numerically identical with a causal power associated with its base property. (p.42)

(CCP) rules out the possibility that the relation between the subvening and the supervening properties is causal because “effects (as property instantiations) generally (and at least sometimes) have causal powers that are distinct from those of their causes (as property instantiations).” (p.42)

The caveats in this definition should already begin to suggest that Wilson’s proposal is inadequate to ground physicalism. In fact, if we regard the causal relation as transitive, as most do for deterministic causation at least, it is not clear to me why it would be wrong to say that causes contain all the causal powers of their effects. Unless Wilson is denying the transitivity of causal powers, which she provides no argument for doing, her statement seems just false. Moreover, if causal powers are universals of some sort, then there will be genuine cases of causal overdetermination which fulfil (CCP). For example, if Smith is shot in the heart simultaneously by two separately fired bullets, then both shots arguably share the causal power (i.e. the numerically identical universal) of killing Smith (and perhaps all others causal powers too). But the relation between these two firings cannot be of the sort that physicalists have in mind for the mental and the physical. One would think then either the relation between these generals must be constrained in some further way or causal powers are to be construed as particulars, perhaps something like tropes. Wilson provides little guidance here. However, I suggest Jaegwon Kim’s (1998) recent writings provide one possible

and plausible way to construe Wilson's (CCP) condition: as a version of functionalism.

On such an account mental properties are to be *defined* as whatever takes an individual from various sensory inputs to various behavioural outputs. According to Kim then what must first be done is to reconfigure mental properties in terms of their causal powers: to functionalise them. The functionalised states are then to be identified with their physical realisers. The mental supervenes on the physical (the brain) because mental states are identical, given their functionalised definition, to certain physical states. Kim's new position is redolent of (in fact identical to, I would say) the old analytic functionalism of the Davids Lewis and Armstrong.

Does this recasting of an old solution, solve all our worries? Well, consider what perplexes Kim most, mental causation. If the mental and physical are identical then there can be no concern that the mental is epiphenomenal in the way that has plagued other supervenience accounts. However, one might be worried by the way we have arrived at the conclusion that the mental and the physical are identical i.e. simply by redefining the mental in terms of the physical. Kim is sensitive to this objection. Mental properties as functional properties are second-order properties defined over the first-order physical base. But are such second-order properties really properties, particularly if we think they may be multiple-realised? Kim thinks not since: "By quantifying over properties, we cannot create new properties, anymore than by quantifying over individuals, we can create a new individual." (p.104) So, second-order, functionalised mental properties are best thought of as not properties at all but "second-order *descriptions* or *designators* or second-order *concepts*" (*ibid.*). Note how similar

this is to Heil's theory of events as tropes. Those events are essentially physical but they may be grouped together in such a way as to fall under the patterns associated with mental states. What mental predicates do then is pick out sets of *physical* tropes, just as what mental concepts do on Kim's view is pick out physical realisers. As with Heil, then it is not really correct to say that mental states cause physical states since there are not *really* any mental properties, just words, concepts, etc. that can be used to pick out physical properties and states that do the causing.

Kim's position also draws on elements of (MS). Before mental concepts can be functionalised, macroscopic objects must be built up out of microscopic objects. This he seems to think is unproblematic. So reduction comes in two stages: building the macro world out of the micro world and then functionalising non-physical parts of the macro world so they can be identified with the macroscopic physical parts. This for Kim is the only way to avoid the problems of causal exclusion argument.

As we can see then Kim's position draws on elements of all three accounts discussed. (MS) is used to build the macrophysical from the microphysical. We are to understand mental properties as a way of classifying physical types as with Heil's and possibly the Macdonalds' theory and that classification proceeds on the basis of identifying causal powers as Wilson suggests. Kim's new position then seems to be the natural end point for the discussion. Interestingly, this has led us back to a kind of reductionism. Kim's position inherits all of the weaknesses of (MS) and it is questionable whether it really gives an account of mental causation. Nevertheless, it is a clearly stated and influential version of the physicalist project; and demands deeper discussion, along with other philosophical attempts to

provide a functionalist-reductionist account of the relation between the mental and the physical.

Chapter 3

Functional Analyses

A return to functionalism and reduction and a move away from supervenience solves the problem of epiphenomenalism which plagued many of the accounts discussed in the previous section. If mental events are just a sub-class of physical events, then there is no question, or at least no special question, of how they can cause certain physical events, given other physicalist assumptions. What we need to consider in this chapter is if functionalism can save physicalism. That is to say, is it all plausible that the mental reduces to the physical in the way that functionalists maintain? In addition to Kim's analytic functionalism, we shall also consider biofunctional versions of the reductionist project. Both kinds of functionalism, it will be argued, fail to make plausible cases for reduction.

3.1.0 Analytic functionalism

Let us first consider Kim's analytic functionalism. The idea here is that mental states can be *defined* in terms of causal dispositions stated in physical terms:

For functional reduction we construe M as a second-order property defined by its causal role – that is, by a causal specification H describing its (typical) causes and effects. So M is now the property of having a property with such-and-such causal potentials, and it turns out that property P is exactly the property that fits the causal specification. And this grounds the identification of M with P. M is the property of having some property that meets the specification H and P is the property that meets H. So M is the property of having P.

But in general the property of having property Q = property Q. It follows then that M is P. (Kim (1998), pp98-99)

What we need to assess then is the plausibility of there being a specification H which singles out each mental state and that H in turn is realised by some physical state Q.

3.1.1 Qualia

One main source of opposition to this account comes from advocates of what is sometimes called qualia, the phenomenal aspects of conscious experience. Frank Jackson (1986) and David Chalmers (1996) have argued that certain thought experiments show that qualia cannot be cashed out in functional terms. According to such thinkers, it is supposed to be possible to imagine someone just like you or me who has all the same beliefs and acts in exactly the same way but who experiences, for example, the colour spectrum as completely inverted. When we see red, he sees purple, when we see orange, he sees blue, etc. So, the argument goes, if that is possible, then our conscious experience of colour cannot be defined functionally since there is no functional difference between me and my colour-inverted double. Moreover, because qualia cannot be functionalised, they are taken by such theorists to be epiphenomenal.

I don't want to discuss qualia and the problems they create in detail for three reasons. First of all because I am not entirely sure what they are. In particular, it is a mystery to me how philosophers like Jackson and Chalmers can form beliefs about qualia, especially their own qualia and yet claim that they are

epiphenomenal.¹ Where do they get these beliefs from? Second, and more importantly in the present context, Kim and some other physicalists are willing to acknowledge that qualia cannot be functionalised. This leaves two options for the physicalist. Either they can go for a type-identity theory with respect to the phenomenal aspects of the mental – Kim adopts this position reluctantly, David Papineau (1998), somewhat more enthusiastically – or they can question the clarity and adequacy of the idea of qualia as formulated by philosophers and opt for a kind of eliminativism with regard to qualia. The third and most important reason why I will not discuss the problems of qualia at any length is that Jackson and Chalmers are both functionalist with regard to other mental states. The arguments regarding qualia are very much a local dispute between philosophers of a largely functionalist/physicalist persuasion. It would be a hollow victory for an anti-physicalist like me to find the only real difficulty for functionalism was created by qualia. For these reasons I want to discuss what appear to me more fundamental problems with functionalism.

3.1.2 Three problems with functionalism

Let us turn our attention to belief, desires and other intentional states. Can they be functionalised and reduced as Kim envisages? There are three general objections to this programme that I will discuss. All are familiar from the literature on functionalism. Nevertheless, they are worth considering in light of the specifically reductionist aims of Kim's functionalist programme.

The first thing to notice is that what we consider mental states to cause are typically actions and these are not individuated in physical terms. For example, if

¹ Chalmers (1996) view seems to involve a mysterious kind of parallelism between functional

I believe that there is beer in the fridge and I desire a beer this will cause me (normally at least) to perform the act of opening the fridge. That would be a typical causal potential of the conjunction of these two mental states. But the effect – the action opening the fridge – is not a physical effect or in any straightforward sense identical to a physical effect. We can see this when we think about the various different ways in which one might open the fridge. I might perform the act just by reaching out with my hand and pulling the handle or wrench it open from the top or maybe open it with my foot if I'm agile enough and my hands are full. All of these seem to be fridge openings, all plausibly the result of the same combination of beliefs and desires but they involve different physical movements. Moreover, the same movement might be a different action. I might raise my arm to hail a taxi or to block someone's passage, to push a door open, etc. So it seems that we do not get the identities Kim's analysis might lead one to expect. Beliefs cause actions and actions do not seem to be identical to *any* particular bodily movement. It looks as though we might have to invoke some kind of supervenience thesis here to link actions with bodily movements; and that begins to look like trouble given the problems we have highlighted with supervenience in the previous chapter. (Not to mention embarrassing for Kim's claim that functionalism is the right way to explain supervenience.)

The second problem relates to what philosophers call *externalism*: the view that intentional states are in part individuated by reference to objects those states concern. The standard line of argument used to motivate externalism involves recourse to some recondite possible worlds reasoning. The story usually goes something like this: imagine there is a world exactly like ours but where

there is a substance that looks like water, has the same macrophysical features as water and happens to be called water too but has a different microconstitution, XYZ, rather than H₂O. Now imagine two physically identical individuals staring at a pool of liquid, refreshing stuff. One thinking that water is wet and his twin on twin Earth thinking that twin-water (a different substance, remember) is wet. Intuitively it is argued, they have different beliefs because their beliefs are *about* different substances. Since they are physically identical there must be more to the content of any belief than simply what goes on in the individual's head.² Such stories are fine as far as they go. However, objections to this kind of possible worlds talk can often obscure what is what is a quite general and readily appreciable point. One which arises in a particular vivid way when one begins to think of mental states in terms of their causal role.

Consider again my desire for a beer and my belief that there is beer in the fridge. It seems reasonable to infer given these intentional states that I will reach out and open the fridge. The relation this belief bears to other intentional states and actions partially defines it, according to the functionalist. If I hold this belief in different circumstances, if, for example, I have a different set of beliefs or desires, or the environment is different, then I will act differently. If the fridge is already open, I won't reach to open it; or if I feel hungover from the night before, I might think better of it; or if I believe that by moaning enough I can induce my girlfriend to get the beer for me, then I might do that instead and so on and so forth. The sum total of all these dispositions is supposed to provide us with the specification H that completely defines the mental state and the state so picked out is, according to Kim, realised by (that is to say, identical to) some physical,

² See Putnam (1975) and Burge (1979) for the glorious details.

presumably neurophysiological, state. So in the case I have described above, Kim is committed to there being some internal physical state which causes the action of opening the fridge (since it is one of its causal outputs, which, as I said, partially defines it). The act of opening the fridge, as we have seen, is not an internal state. It clearly involves reference to an object outwith my own body, namely the fridge. Given the functionalist story Kim wants to tell, if mental states like beliefs are partly defined in terms of the actions they produce, then if those actions are individuated by objects external to me (which, as we have seen they are) then so too must the beliefs.

So an intuitive understanding of functionalism leads to a form of externalism; and that looks like bad news for Kim since if mental states are *identical* to anything one would think it would be brain states. But if mental states are what they are in virtue of factors external to any thinker this can't be right.

The third problem is not everything that might be caused by some mental state should be essential to that mental state being what it is. For example, my belief that there is beer in the fridge might cause my heart to beat faster with excitement. It might, on the other hand, elicit no such reaction in a more moderate person. According to the simplest analysis of functionalism in that case we should hold different beliefs. The specification, H, associated with my belief-that-beer-is-in-the-fridge has different causal potentials to the moderate person's belief-that-beer-is-in-the-fridge. But that is, I take it, absurd; such effects should not matter to the individuation of belief. Kim is clearly aware of some sort of problem like this since he limits the specification H (in parenthetical apology) to *typical* causes and effects. The problem for the functionalist is to say what typical means here in a principled way that does not undermine his reductionist programme.

3.1.3 Narrow and broad contents

One possible solution to some of these problems is offered by invoking a distinction between narrow and broad content. Advocates of such a distinction accept the general externalist point that the content (of some) intentional states is dependent on facts outside the head. However, they will claim that the right way to think of what is going on here requires a two-part story of content. Part of what it is to believe that beer is in the fridge is to do with something internal to the individual who holds such a belief and part is to do with the environment of the individual. So, roughly speaking, there is a story to tell about what is going on in the individual's head which when combined with the right environment will constitute the belief that beer is in the fridge. The narrow-content is that internal state. It also follows on this story that if the same internal state were combined with a different environment, it might be part of a different belief. For example, if it were an environment where there was no beer but only a similar drink, queer, let's call it, brewed from quops and quarley but otherwise indistinguishable, then the same narrow content would be part of a different wide belief – namely the belief that there is queer in the fridge. Such theories of content are often called 'two-level' semantic theories since part of the content is fixed by the narrow state and part by the environment and context; and it is the narrow part that is to be functionalised and reduced.³

Clearly this seems to represent the beginning of an answer to the second objection I raised. It raises hope of a solution to the first objection too. What makes the action, the action it is, is the narrow state combined with the right

environment. So if the idea of narrow content defined in terms of physical inputs and outputs can be made good, then we would have an explanation for the relation between intentionally described behaviour and bodily movements. That is to say, one could argue given John has narrow content X, that will cause movements A, B, and C which when combined with the particular environment constitutes some action or other. In the abstract this looks good. What we need now is a defence of the notion of narrow content that could do all this and be apt for Kim-style reduction.

Narrow content entered the philosophical lexicon at the same time as the twin Earth thought experiments and as my sketch of the idea should illustrate, the sense of the concept is strongly dependent on that counterfactual structure. Away from the metaphysically rarefied atmosphere of possible worlds reasoning, the general notion of narrow content seems pretty hard to motivate. In particular, it seems difficult to articulate an interesting account of the identity conditions for narrow contents. Consider two obvious proposals. One might say that X and Y have the same narrow contents when they have the same beliefs. That would seem to make the idea of narrow content entirely superfluous and conceptually dependent upon normal (that is wide) content. One might wonder what work such a concept is supposed to be doing. Twin Earth thought experiments, it might be maintained, could come to the rescue. The definition might run as follows: two creatures have the same narrow content when they are in exactly the same internal physical state. The advantage of this definition is that it allows for variation in wide contents; narrow contents are liberated from their dependence on ordinary contents. However, as a serious theory of psychological states it does not look any

³ The narrow content is often thought to play a role like Fregean sense: distinguishing belief states

better than the first proposal. No two *actual* creatures ever are in exactly the same internal physical state. This definition renders narrow content useless for any real psychology. A serious advocate of narrow content needs some means of defining his notion which is independent of wide content and mere physical identity, if it is to be of any theoretical interest at all.

It may seem that functionalists have a ready-made solution to the problem; defining a state by its functional role would seem to require neither wide-content nor physical identity. However, as the arguments I have outlined above demonstrate, if functional role is related to action, then that has unavoidable externalist consequences for beliefs and other intentional states. What the functionalist-reductionist requires is a plausible functionalist account of content that operates with only narrow states.

Conceptual role⁴ semantics (hereafter, CRS), as championed by Ned Block (1986) and Hartry Field (1978), advertises itself as fitting this very bill. According to CRS the narrow part of the content of a belief is given by its conceptual role, which is just a more rationalistic version of the functionalist idea that it is given by its causal role. In fact, we can see it as the beginning of the answer to third problem I raised for functionalism. The typical connections Kim needs are the inferential ones. While, like other two-level theories, environment and context fix reference and truth. So we have a possible answer to all three problems that I posed for functionalism. A CRSist can claim that an account of narrow content can answer the first two problems and construing the appropriate or typical relation between mental states in terms of inferences can answer problem three.

like *Muhammad Ali is the greatest boxer ever* and *Cassius Clay is the greatest boxer ever* which are indistinguishable in terms of their broad contents (i.e. their reference).

⁴ People, perhaps more accurately, also talk of *inferential roles*.

The idea that the conceptual role is in some way important to the meaning of a term has some intuitive appeal. Block takes his view to be a more scientific and explicit rendering of Wittgenstein's slogan that "meaning is use" and it is not hard to provide a general sketch of the idea that suggests it has a certain plausibility. For example, given that someone holds some belief, call it X, one would expect that person to be able to exhibit certain patterns of inference which demonstrates this fact. To take an elementary example, it might be considered constitutive of having the belief P&Q that one is able to infer P. The CRS theorist claims either that there are similar inferences for all such beliefs and the set of these will individuate each (narrow) state or that there exists a basic stock of beliefs which can be defined this way from which all one's other beliefs may be constructed.

The most obvious (and extremely serious) problem with CRS is that there just do not seem to be the sorts of conceptual relations between beliefs and other mental states that would fix content in the desired way. Consider some of the well-worn examples from the philosophy of science. It is well known, for example, that people possessed different beliefs about the sun than we do now. So they would be willing to make inferences from the belief that they are seeing the sun which we would not and vice versa. (For example, they would infer that they are seeing a planet, we would infer that we are seeing a star.) But, intuitively speaking at least, we would not say that we possessed different beliefs. However that appears to be precisely what the CRS theory commits one to.

In fact that is very much the tip of the iceberg. CRS threatens to make it impossible to disagree with people or change your mind. Consider my belief that beer is in front of me. I might be inclined to infer from *that is beer* to *that is*

refreshing. Such an inferential connection would partly constitute the belief *that is beer*.⁵ But what if I have a tea-total friend who does not infer from *that is beer* to *that is refreshing*. We have different specifications H for the belief *that is beer*, so, according to the functionalist, we have different beliefs and thus we cannot disagree because what I mean by beer and he means by beer are different. Say I change my mind. I become dehydrated once too often after consuming too much alcohol. I no longer believe beer is refreshing but then my own inferential connections change. So what I mean by beer is refreshing now is different from what I meant earlier by beer is refreshing. The belief has different inferential connections. So, I can't change my mind because any apparent change in mind must cause a change in content of the belief.

It is obviously open to the CRS theorist to respond by claiming that only a certain stock of inferences are necessary for sameness of belief, not an absolute match. If this sort of response is to be made credible, then CRSists need to be able to provide principled reasons for saying what is and what is not content constitutive. That is to say, they still need to provide some more specific response to problem three. No-one has actually done this and I think if we attend to some facts from everyday psychology it seems unlikely that it could be done. We are all familiar with the psychological phenomena of being shown that our beliefs have some surprising or unwanted consequence or failing to recognise something which we later take to be an obvious consequence of our views. Philosophers should be more familiar with this than anyone else. The industry of filling philosophy journals is partly driven by the fact that we often miss the conceptual or logical consequences of some of our beliefs that other people spot. The natural

⁵ In some derivative sense, it would also partly constitute the meaning of the terms that compose to

way to describe such situations is to say that we held some belief but we failed to recognise that it had a certain consequence. Such phenomena show how difficult the task before CRSists is. Even inferences that we often take to be obvious are not necessary for holding some belief. So it seems very unlikely the CRS theorists could provide a story of content-constitutive inferences which would be at once rich enough to individuate content and at the same time accommodate the obvious fact that we often miss or fail to notice consequences of our beliefs.⁶

Block's(1996) actual response to the problem of the instability of content in light of changing of beliefs and inferential connections has been to claim that the two-level nature of his version of CRS could come to the rescue. The broad part of content acts as some kind of constraint on the variations in narrow content. Beliefs about beer are still beliefs about beer despite changes in inferential connections because the environment fixes them as such. How this is supposed to work is utterly unclear but it does point to a problem with CRS that we have up until now been overlooking.

So far we have been taking the idea of inference for granted but in the context of narrow content theories, the very idea of conceptual role as used here is problematic. When we consider relations between ordinary beliefs, the idea of conceptual role has at least some intuitive plausibility but what on earth (or twin Earth, for that matter) are narrow conceptual roles? How are we to understand the notion of inference when we are only dealing with narrow contents? An objection that Jerry Fodor and Ernie Lepore (1991) have raised becomes particularly pertinent once we recognise this: what guarantees that the two elements of any two-level semantic theory like CRS will work in tandem? What is to stop the

make the belief. That is to say, the meaning of the term beer is to be explicated in terms of all the

narrow conceptual role individuating a belief content as *water is wet*, the truth conditions for which are given by *4 is a prime number*, for example? When we consider ordinary conceptual connections between beliefs this objection can seem odd. It is surely evident that the role our beliefs play do indeed cohere with their reference. Yet we can only be so sanguine concerning the apparent relations between the putative parts of the semantics because when considering the conceptual role we already have in mind ordinary (i.e. wide or truth-functional) beliefs and other mental states.

These considerations show the poverty of the CRS theorist response to the problems with functionalism I have outlined. An appeal to narrow content was supposed to help answer the first and second difficulties that we encountered. If that is to work then we need some clear account of narrow content and how it combines with the environment to form wide contents that plausibly can be said to be the causes of actions. But we have just seen CRS theorist have no such account. We don't know how to answer problems one and two because we don't have an account of narrow content. We don't have an account of narrow content because we don't know what is meant by narrow inferences. We don't know what is meant by narrow inferences because the idea of inference seems to involve wide or truth functional contents. And the CRS theorist has given us no story or theory as to how narrow inferences, whatever they are, are supposed to be linked to normal inferences. Far from solving the problems of mental contents in a functionalist manner, CRS merely makes them more vivid. We still have no story which tells us what narrow contents are or how they connect with wide contents.⁷

beliefs in which it appears.

⁶ This point is also made by Putnam (1999), p.117

⁷ One possible response (see Carruthers and Botterill (2000)) is to suggest that some new psychology will make sense of the relevant connections for us. That is to say, there will be a

One final way of overcoming some of the problems with functionalism that we have raised is worth considering. We might be able to answer problem three and avoid some of the absurdities of CRS by including not only actual inferential or causal connections but also counterfactuals ones (see for example Carruthers (1996), ch.4). So although the actual conceptual or causal role of my beliefs about the sun are different from someone in the past, if the other beliefs they held were the same, then they would draw the same inferences and act in the same way as me. This looks like it might solve many of the problems that have confronted functionalism and it is, perhaps, suggested by Kim's talk of causal *potentials*. However, it is a non-starter.

First one should notice that a specification of all the potential causes a belief might be involved could be endless given the infinite number of possible combinations of beliefs and other intentional states there could be. The specification H would be impossible for anyone to write down or conceptualise; and thus any actual reduction would be impossible. Moreover, given the possibility of such an open-ended disposition, one might start to doubt Kim's claim that there could be one simple neurophysiological state that could be implicated in such a variety of possible effects. It just seems implausible that there is one physical state that could be causally implicated in the vast variety of things one might do given that one had some particular mental state. Just consider the

psychological theory that picks out narrow states, the laws of which will relate one state to another in a way which is consistent with what are intuitively the right inferential connections. In fact, it might do better than that: it might correct our intuitions about inferential connections between beliefs and so improve upon our commonsense psychology. This suggestion can be summarised (in only moderately tendentious terms) as follows: there is a something or other (narrow content), governed by some laws or others (new psychology) which will solve all our problems (hooray). But of course no-one has any idea what this sort of new psychology will be like and when and how it would be applied or tested. This is all pie-in-the-sky. Impossible to discuss because there is no actual theory that can be assessed. Furthermore it is inconsistent with analytic functionalism since it would need to presuppose some prior understanding of narrow contents on the basis of which

endless way one might further supplement the possible actions I might take given I believe beer is in the fridge. Perhaps this does not look like such a huge problem. Can't the functionalist invoke multiple-realizability here? So there might be a range of causal specifications, H, H', H'' ..., which each pick out a subset of the dispositions associated with any given belief. Each of these specifications in turn would be identical to a different physical state. But that would be a disaster for the functionalist project. What would these causal specifications have in common by physicalist lights; on what basis should we regard them as the *same* belief? Multiple-realizability of this sort excludes the possibility that they are physically or indeed even functionally identical.⁸ What else could be said – that the mental states supervene on the physical basis? We know that won't work.

Such arguments are as fine as far as they go and help to highlight the implausibility of the functionalist claim. However, Kim and others may stubbornly insist that there is in fact one neurophysiological realiser that has all the dispositions we might associate with a given mental state and that reduction in principle is thus not ruled out. The real problem, however, with the suggestion that we try to individuate functional items by including counterfactual connections is that what is and is not relevant to the counterfactual is most plausibly decided by a prior grasp of the semantic content and what it is reasonable to think and do given such mental states. Consider the following story: because someone has the belief that seagulls have gold in their stomachs, and they have a desire for fortune, we might think that they would try to hunt down and disembowel seagulls. On the other hand, they might have further beliefs about gulls, say that they are in some way sacred which prevents them from doing this, however much we may imagine

psychological laws might be investigated. It would thus make the connection between narrow

they secretly desire the gold they believe is there. We can tell stories like this, even if as a matter of fact no-one has ever held such beliefs because we understand such beliefs to be about the world; that is to have a certain content. Then given somebody thinks the world is some way we can further imagine how he or she might *reasonably* be expected to act.

It is always tempting to think that there must be a functionalist story to tell about mental states because we can quite easily construct stories like this to explain why someone with such and such mental states did what he did. And of course, we can construct lots and lots of stories about what people might do given that they have or might have certain intentional states; such is the basis of our ability to tell stories at all. It is tempting, therefore, I suppose to think that the sum of all these constructions provides the functional specification that Kim is after. But that overlooks, of course, how it is that we are able to construct these counterfactual scenarios. We can do it because we have a grasp of the contents of the agent's mental states and then given also the idea that the world is some way or another, we can infer what might *reasonably* happen, as we do in the case of man who believes that there is gold in the stomach of seagulls. To make sense of the counterfactual claim in the required reductionist sense we would need a criterion of identity of the state which is independent of our understanding of the belief state *qua* belief state. But we do not have that. Our understanding here is entirely top down. So although it is always apparently possible to save functionalism by appeal to counterfactuals connections of the form we have discussed above, the result is a Pyrrhic victory for the reductionist. The resulting theory depends upon a prior understanding of the content of the beliefs and what it

states and inference an empirical one, not a definitional one.

might be *reasonable* to think and do given such beliefs. It is in virtue of such considerations that we can decide what is and what is not relevant. Since this already assumes the content is what it is and because there is no functionalist story to tell us how to think about this notion of reasonableness or rationality (that is to say, still no answer to problem three), there is no reduction here.

To put it succinctly, the problem with CRS is that it does not exist.⁹ It fails to answer problems one and two because it fails to provide an adequate account of narrow content and how it is connected to truth-functional content. It also fails to answer problem three. When we construe CRS in terms of the actual inferences that any individual makes, CRS leads one to the absurd conclusion that we can never disagree with anyone or change our minds. When we consider counterfactual connections as well, then we can avoid this problem but only at the cost of giving up on reduction. Deciding which counterfactual are relevant requires a prior grasp of the mental contents and intuitions about what it is reasonable to do given such mental states. A reductionist treatment here would have to analyse away or otherwise explain this idea of what it is reasonable to do. No CRS theorist has any suggestion how this might be done. Moreover, reflecting on the kinds of inferences and relations between mental states that we have (and sometimes do not have) just makes the whole idea of CRS implausible. For almost any mental state, it is surely possible to construct some counter-example in which we attribute say belief X to so-and-so but not one of the supposed meaning constitutive inferences.

⁹ Similar arguments can be found in Prades and Corbi (2000) and Ben-Yami (1999).

3.1.4 Eliminativism

A more radical solution to the three problems I outlined at the beginning of this chapter is possible. Recall that Kim does not think that there are *really* any functional *properties*, merely functional and therefore mental *concepts*. Strictly speaking then one might want to interpret Kim's position as a version of eliminativism.

Now this could be a quick way of disposing of most of the objections I have raised. If there aren't really any belief or desires then there is no reason to agonise over the exact set of inferences or causal connections which defines any mental state. Nor is there any reason to be too concerned about how it is that we arrive at that conclusion since no such decision would carry any metaphysical weight; such states would not have to refer to anything. Rather we would just have certain concepts that could be used to help us get by in the world. I do not suggest that Kim's own quasi-eliminativist views are as radical as this. He does not arrive at his eliminativist conclusions because he believes that psychology is at best of instrumental value; rather eliminativism is forced on him because of the physicalist idea that the only *real* properties are physical properties. Nevertheless, once the idea of eliminativism has been introduced, it is tempting to use the general idea as a way of overcoming the problems that functionalism has encountered.

I don't think this can possibly be the solution functionalists are after. Consider the claim shared by Kim and more radical eliminativists that there are not really any mental properties, only mental concepts. What on this account are concepts? On the radical eliminativist line it does not seem concept talk

⁹ As Paul Boghossian says, "Inferential role semantics is no more than a twinkle in the eye of its

corresponds to anything. There can be no concepts, any more than there can be beliefs because concepts do not belong to the basic physical ontology. But then how are we even to understand the eliminativists claim; a claim which clearly employs concepts? In short, an eliminativist uses certain concepts to express the belief that there are no beliefs or concepts. I can make no sense of this claim.

I think an analogous problem arises for Kim, even given his weaker eliminativist line. It is not entirely clear what Kim means by concepts and how they would fit into a physical ontology. Perhaps there is some reasonable answer that can be given here in terms of the causal potentials of agents. So you might define a concept by reference to all the mental states in which they can figure, which are in turn defined by their causal potentials. But Kim's account still has features which trouble me. To *understand* a theory like Kim's we need to be able to grasp the concepts involved – both mental and physical. We need to understand how such concepts can be related to form complex concepts – like physicalism and anti-physicalism; and then Kim and I need to understand that we disagree about the *truth* of sentences or propositions employed using such concepts. If all there is to such a disagreement is supposed to be given (or analysed away) by a set of causal dispositions, then the sense in which we might be said to *disagree* seems to be difficult to recover. What I take it Kim and I are disagreeing about is *how things are*, not which causal dispositions one should have. However, in the context of Kim's ideas of what mental states are, it seems that one cannot recover that sense of disagreement. Having the belief that physicalism is the true ontology and having the belief that physicalism does not make any sense turn out just to be ways to *designate* certain physical states. Like the more radical eliminativist,

advocates". And quite probably less.

accepting Kim's position makes it difficult to make sense of the questions it was designed to address; questions about how the world is.

3.1.5 Why functionalism?

Kim-style functionalism suffers from a poverty of explanation. First, it requires a we-know-not-what relation between the microphysical and the macrophysical to guarantee (MS). Then it requires at very best a we-know-not-what theory of psychology involving narrow contents (whatever they are) that can be functionalised and reduced to the macrophysical. Kim's return to reductionism has brought little in the way of enlightenment as to what the relation between the mental and the non-mental might be.

What is perhaps most perplexing is that having once espoused anti-reductionist views and thought, perhaps as many did that this suited the phenomenology of the sciences better, that Kim should feel compelled to return to hopeful reductionism in order to avoid the problems of the causal exclusion argument. A position which marks out little positive but makes a few vague promises that things will work out in the end. I find it difficult to see what the attraction of such a view is supposed to be. I find it difficult to believe that Kim finds it attractive given remarks like the following:

I don't think it is good philosophy to say as some materialists used to say, "why can't we just say they [the mental and the physical] are the same? Give me good reasons why they shouldn't be the same." I think that we must try to provide positive reasons for saying that things that appear to be distinct are in fact one and the same. (p.98)

Likewise, I don't think it is good philosophy just to say something which does not look like it is functionalisable, is. But Kim, appears to do no more than that:

There has been much skepticism about the viability of a functionalist account of intentionality ...However, like many others, I remain unconvinced by these arguments; I don't see any principled obstacles to a functional account of the intentionality. (p.101)

What positive reasons are there supposed to be for believing functionalisation is possible? Only "that it seems inconceivable that a possible world exists that is an exact physical replica of this world but lacking wholly in intentionality." (p.101) But what on earth are such appeals to intuitive conceivability supposed to show? That there must be a reductionist (functionalist) story to tell about intentionality in terms of physics? I find it inconceivable that anything could be coloured without occupying some space. Does this mean there must be a reductionist story to tell about colour in terms of extension? Obviously not. There is no reason to believe that the mental can be suitably functionalised and, *pace* Kim, I would suggest the problems adumbrated above should make us sceptical that there can be any such account.

Let's move on. Analytic functionalism looks wholly implausible account of anyone's mental life. A better alternative, more obviously naturalistic in spirit, which might still be broadly considered to be physicalist and functionalist, is provided by what is sometimes called teleosemantics. The general idea here is that

the notion of functionality found in biology may be used to effect a reduction of psychology. The best (i.e. the most entertaining) way to discuss teleosemantics' reductionist claims is in terms of a well-known thought experiment.

3.2.0 The man from nowhere

Suppose lightning strikes a dead tree in a swamp; I am standing nearby. My body is reduced to its elements, while entirely by coincidence (and out of different molecules) the tree is turned into my physical replica. My replica, The Swampman... moves into my house and seems to write articles on radical interpretation. No one can tell the difference [between him and me]. (Davidson, 1986)

But there is, of course, a difference between Davidson and the imagined Swampman; a difference in causal origin. Is that an important difference? Philosophers are divided. Some say the peculiarities of Swampman's origins mean he cannot have beliefs or desires; he cannot, in short, have a mind. On the other hand, some philosophers believe that that it is unreasonable to say that something that looks like a philosopher, walks like a philosopher and most importantly talks like a philosopher, is not a philosopher and therefore (though perhaps this step is questionable) a thinker. Advocates of the mindless view fall into two broad categories: what I shall call historical externalists and teleosemanticists. The historical externalist thinks that causal history of the individual is vital to the meaning of the words he uses and the thoughts he has. He believes that the process of language learning is somehow essential in making

someone a thinker. This is certainly Davidson's view. The teleosemanticists, although they may agree that learning is important in determining semantic content, think that it is only part of the story. Prior to the correct learning process, thinking beings must have the correct phylogeny. That is, they must have an evolutionary history from which emerged a certain set of characteristics that were selected for the function of thinking¹⁰. No evolution, no thought, no mind. Ruth Millikan is undoubtedly the leading advocate of this bio-functional approach, among her acolytes we may count David Papineau, Karen Neander and sometimes Daniel Dennett.

My concern here is not really to decide the question: does Swampman think. Rather, I am interested in using Davidson's famous thought experiment as a means of expounding and criticising teleosemantics as a reductionist account of content. But before I look at the theory in any detail, I feel I need to say a few words in support of the value of Swampmen (whether thinking or non-thinking).

3.2.1 Thought experiments, naturalism and intuitions

Almost all philosophers who attach themselves to teleosemantic views proudly associate themselves with a strong version of the naturalist doctrine. With this kind of naturalism tends to come a certain hostility to what are perceived as traditional ways of doing philosophy; ways of philosophising that, for example, rely heavily on *a priori* arguments, logical possibilities and our supposed intuitions. Many such naturalists thus quickly and airily dismiss Swampman as an example of this rather limited and not very insightful way of thinking. Dennett expresses this bristling impatience most vividly:

¹⁰ When I say thinking, throughout this section it is meant to imply thought which involves

Does Swampman have thoughts and use language, or not? Is a cow-shark a shark? It swims like a shark and mates successfully with other sharks. Oh but didn't I tell you? It is atom for atom indistinguishable from a shark except that it has cow-DNA in all its cells. Impossible? Not *logically* impossible (say the philosophers). Just so obviously impossible as to render further discussion unnecessary. (Dennett 1996)

Similar, although less aggressively stated, thoughts can be found in Millikan (1996) and Papineau (1996). Beyond the rhetoric though what is supposed to be so wrong with Swampman. Is it that it is only a thought experiment, an event never likely to happen in this world? It cannot be that simple. All philosophers, even Dennett, employ thought experiments. More importantly, many scientists employ thought experiments. For such naturalist philosophers, sense and nonsense are merely mispronunciations of science and non-science; if scientists do it, then it must be all right. No, it cannot simply be that it is a thought experiment, which so offends Dennett, it must be a *bad* thought experiment.

What then differentiates a good thought experiment from a bad one? Dennett seems to be implying that Swampman is *physically* impossible; that is to say, that imagining such an occurrence contradicts the laws of physics. If this is so, it is obvious why, by naturalist lights, Swampmen are thoroughly disreputable. We should only be interested in the world as described by science not by various crackpot philosophers. But is it really the case that physics rules out Swampmen?

There might be certain reasons to think so. Certainly it cannot be the case that the right sorts of *chemicals* are just flung together at random. The macromolecules that are the building blocks of life can only be formed by a series of quite specific complex chemical reactions. Random collisions of even these molecules will not create Swampman, just more swamp slush.¹¹ However, perhaps we can by-pass the chemical domain altogether and create Swampman simply from the spontaneous coalescence of fundamental physical particles; or at the very least we might say that *if* one is a physicalist, this should be entertained as a possibility.

There might be reasons to doubt this too. The tremendous energy required to do this may make any such structure unstable. Of course, this is not to say it is physically impossible as Dennett insists. Rather, we are simply uncertain one way or the other.¹²

I think, though, we can retell Swampman stories in a way that maybe makes them more acceptable to naturalists. After all, Davidson's particular description is not important. What matters is one or both of two things. First that this being lacks an individual history and second that it lacks an ancestral history. There are a number of different ways one may 'create' beings with such features. For instance, one might imagine that in the future we are able to create a replicator of some sort. That is to say, we might invent some sophisticated bit of technology which can scan a person's physical structure and then reproduce that structure from different elements. Is this physically possible? I have no idea but enough naturalists discuss such examples to make me confident that they have no special philosophical problem with the concept. Of course, the reason that many

¹¹ Millikan (1996) and Neander (1996) both make this point against the possibility of Swampmen.

¹² Of course, if we were to think about what is going on in terms of statistical mechanics, then there is no reason to suppose Swampman should cause any special difficulty. There are many different paths particles can take to end up in the same position, with the same velocity.

philosophers, in particular those hostile to Swampman, discuss the possibility of replication by such machines is to distinguish it from the spontaneous coalescence of Swampman. Replication by a replicator is non-accidental; and this fact, so it is argued, allows one to think of the replicated being as having both a history and an ancestry. Because the replicated being is an intentionally created copy, it can, so to speak, inherit all the features of the being that it is a copy of. Even if we grant this is true, replication by a replicator might still be accidental in at least two ways.

Let us imagine I have been given a replicator for Christmas. I have never seen one before, and so do not recognise what it is. Suppose further that the generous individual who has provided me with this gift has neglected to enclose any instructions. Somewhat puzzled by my piece of gadgetry, I start to fiddle with it. Unbeknown to myself I flick the thing on to replicate me at the same time as I accidentally electrocute myself on the Christmas lights. At the moment I die a replica of me is produced by the machine. However, no one intended to do this. I, after all, did not even know what the machine was supposed to do. So my double is accidental, like Swampman.

It might be argued that my intentions are unimportant here. It is the fact that the machine has been reproduced in order to replicate that matters. So anything replicated by it does, so it happens, inherit all the functional attributes of the original. However, one can easily construct another tale to counter this objection. We are back in the swamp and Davidson, for it is he, is not there just for a pleasant stroll. No, it transpires that Davidson has been sent there to construct a secret military device. He has been working on this weapon for many weeks now but with little success. The weapon is supposed to reduce human

individuals to their constituent elements but what he has actually and quite inadvertently invented is a replicator. Davidson turns it on and is simulataneously destroyed by a lightning bolt and replicated by his machine. The function the machine performs is accidental, therefore Davidson's replica is, it seems to me, truly a Swampman.

If one is unhappy with the idea of replicators, another tale might be told in which we might find a creature without ancestors. The teleosemanticist, as I said earlier, dislikes Swampman because of his lack of pedigree, the historical externalist dislikes him because of his lack of education. One way to make this contrast clear is to consider a Swampbaby. The teleosemanticist will claim that this baby is a non-thinker, and will always be a non-thinker. The historical externalist will claim on the other hand that Swampbaby is no different from any other baby, he just needs time to be educated. I can think of at least one other, admittedly far fetched but possible in a respectably naturalistic understanding of the term, story that would produce a kind of Swampbaby. The tale involves time travel so let me defend this idea first.

If we take general relativity seriously there is no reason to suppose time travel is impossible. Indeed there are certain space-time structures, Gödel space-time structures, in which time travel would be no more problematic than space travel. Let us imagine that we are in such world. Within this world there exists a twin boy and girl. They have been separated at birth. One day each independently decides to go back in time – till approximately a year before they were born. There they meet up, fall in love and soon have twins of their own. But their children are actually just their young selves; they are each their father and

mother.¹³ Now this story is more than merely a mix of science fiction and Greek tragedy – it is of some philosophical importance. The result is two humans with no ancestry whatsoever. So on the teleosemantic view neither of these individuals can be described as a thinker: they are both Swampbabies. So even if you eschew the idea of replicators, you can still obtain Swampman-style examples through a normal process of human¹⁴ sexual reproduction.

Of course, one might not be persuaded that any of these examples is more reputable than the original thought experiment. Replicators and time travel might be too far fetched for some. I must admit that replicator talk in particular makes me a little metaphysically queasy. Nevertheless, this much does hold true, I think: if you are a physicalist, there seems little or no reason to rule out Swampmen; and the time-travel story gives you a way of understanding how it might, in fact, be possible. Since the target of this work is physicalism that is enough for me.

3.2.2 The teleosemantic view

Some details need to be filled in to see exactly why on the teleosemantic view Swampmen just ain't thinking men. The most elaborately worked out of all the teleosemantic theories is Millikan's, so I shall concentrate on her position in the following and demonstrate that once the full details are worked out it leads to an unexpected result.

The keystone in the arch of Millikan's philosophy is her account of "proper function". She defines the phrase thus:

¹³ I have to thank Phil Dowe for a talk he gave on cloning using time-travel for giving me the idea of swampbaby.

¹⁴ One should note that strictly speaking on the teleosemantic view it would be wrong to call either individual human. I shall explain and criticise the reasons for this in the following section.

[F]or an item A to have a function F as a “proper function” it is necessary (and close to sufficient) that one of these two conditions should hold. (1) A originated as a reproduction (to give one example as a copy or a copy of a copy) of some prior item or items that, *due* in part to possession of the properties reproduced, have actually performed F in the past, and A exists because (causally historically because) of this or these performances. (2) A originated as the product of some prior device that, given its circumstances, had performed F as a proper function and, that under those circumstances, normally causes F to be performed by means of producing an item like A. (Millikan, 1993, p.14)

The disjunctive definition corresponds to two ways in which an item obtains its proper function either directly by copying (or reproduction, say) of ancestors. Or, as apparently in the case of human artefacts, the function is derived from the device that produces it. The great advantage of this account of functionality is that it provides a naturalistic understanding of the normativity of functional properties and hence of malfunction. This is best seen with an example.

Take a paradigm case of a proper function: a body organ, let us say the heart. We can say that the proper function of the heart is to pump blood round the body because the ancestors of chordate animals had hearts which did, in fact, pump blood and because of this they survived and reproduced. Nevertheless certain animals related to the appropriate chordate ancestors may have hearts that failed to pump blood. These hearts would be malfunctioning. On Millikan’s account it is easy to see what this malfunctioning amounts to. The historical

element, that is the selection history, in the definition of a proper function provides a normative standard by which we can judge whether an item is performing as it should. In other words, when we consider the question of whether a certain item is functioning or malfunctioning we are, in effect, asking the question: does it perform the function it was *selected* for? Millikan claims that if one attempted to define functionality ahistorically¹⁵, for example in terms of present dispositions, one would lose this normative element; malfunctioning would simply collapse into having no function.

This clearly appears to be an improvement on Kim-style analytic functionalism. Recall that one of the serious difficulties for analytic functionalism was providing a principled discrimination between the sorts of causes and effects that were essential to the functional state being what it was and those that were merely accidental. Such a distinction falls straight out of a teleosemantic analysis: the relevant causes of any proper functioning item are the ones it was selected to perform. It is a proper function of the heart to pump blood round the body but not to make a thumping noise that can be detected by a stethoscope. The former is a selected for function, the latter is not.

Notice as well that this kind of normal functioning does not equate with the statistically average functions of an item. The normal function of sperm is to fertilise an ovum since that is what it has evolved to do, even though very few sperm actually perform this function. The teleosemanticist's project is to take this general account of proper function and apply it to representational states.

With the help of the technical notion of a proper function, we can formulate Millikan's objection to Swampman more precisely. First, although it is

¹⁵ As for example Bigelow and Pargetter (1987) have attempted to do.

clear Swampman is a copy of Davidson (an exact copy of Davidson) this is only accidentally the case. It is not true that Swampman is a reproduction of Davidson. Swampman is not a reproduction of anything; he is a newly created being with no ancestral history. It is pretty clear then that given he has no ancestral history he can have no proper functions in the sense implied by the first half of the disjunctive definition. His heart cannot be said to have the function of pumping blood. His eyes cannot be said to have the functioning of seeing. And his brain cannot be said to have the function of thinking. All of his apparent functional traits are just that – apparent.

A little bit more I think though needs to be said about what the criteria are for ascribing proper functions to any item for there is an apparent ambiguity in Millikan's account. An interesting illustration of this ambiguity might be to take the book of Genesis as more accurate than commonly supposed. Imagine that the human race began with Adam and Eve. But they were not created by God, they were in reality swamp people. On Millikan's account it is unambiguously the case that neither SwampAdam nor SwampEve has proper functions. But what about Cain and Abel? Since they both have two ancestors is that sufficient to ground the ascription of proper functions to them? If not, how many generations do you have to go through before you can? Could we truly say that Noah's heart had the function of pumping blood? Or that David was a thinking man? Here is what Millikan says on the matter:

The functional trait must be one that is there in contrast to the others that are not there, because of historic difference in the results of these alternative traits. It must be tied to genetic materials that were selected

from amongst a larger pool of such materials because of their relative advantageousness... Graphically, whether my shoulders have as a biological function to hold up my clothes depends on...whether there were once shoulderless people who died out because they had nothing to hang their clothes on. (Millikan (1993), p.38)

What is important then in the ascription of a proper function is a relevant contrast class of beings that lacked a particular disposition (but therefore presumably had different dispositions) to compare against the class of beings we are interested in. That any function truly is a proper function in this sense will therefore be quite difficult to prove. It will require a detailed knowledge of populations and how different dispositions that individuals possess within that population affect their abilities to reproduce. It also follows then, and I don't suggest that Millikan would deny this, that many, if not most, of the functions teleosemanticists talk of as examples of proper functions are not known to be so in the sense just described. (I do not suggest in any way that this vitiates the teleosemanticist programme or that there may not be good reasons to think that only an adaptionist explanation could account for certain structures. For instance, no other explanation it might be thought could account for the complexity and arrangement of biological systems. Although, of course in the case of Swampman such an explanation would beg the question.¹⁶)

¹⁶ There might be other problems though related to the unit of selection problem. It is usually assumed that a function is useful, which is why it is reproduced. However, if we believe something like the selfish gene model then features may arise that fall under Millikan's account of proper function without being beneficial to the actual animal. See Manning (1997) for a discussion of meiotic drive as a possible example of this. This is presumably why Millikan claims her definition is *close* to sufficient.

So to summarise: the ascription of functions to the descendants of SwampAdam and SwampEve does not depend upon some arbitrarily chosen number of generations being reproduced but upon the contingent existence of members of the human population that lacked certain causal dispositions; and that those creatures did less well, that is produced fewer offspring, than those who had those dispositions.

The explanation of Millikan's account provided thus far only covers one half of her disjunctive definition. To understand the second part of the definition we need to introduce what Millikan elsewhere refers to as a *derived proper function*. Unlike the first account these proper functions do not have their functions in virtue of a selection history but by being produced by something that has a proper function. We need some extended definition of proper function like this if we are hope to explain all mental states in biofunctional terms. After all, it is very clear that some kinds of mental states involve entirely novel content and thus could not be covered by the first half of Millikan's definition. For example, nobody thought of space and time, like Einstein before Einstein. The second half of the disjunctive definition is supposed to help here.

Consider a chameleon that is able to turn its skin a shade of brown that no chameleon has ever produced before in order to match its surroundings. We may still say that the proper function of this shade of brown is to make it invisible to predators, even though by hypothesis no chameleon has ever been that colour before. The reason we may do this is that the novel shade of brown is produced by some mechanism in the chameleon which has the normal selectionist-functional explanation for its existence. The proper function of the colour is *derived* from the function of the colour-selection mechanism.

Again, *prima facie*, this definition seems to eliminate Swampman from being considered a thinker. Swampman is certainly not *produced* by anything which itself has an evolutionary history. So it would appear there is no way to construe Swampman as a thinker. However, Millikan elaborates her account of derived proper function to deal with some other examples. For example, although the proper function of a turtle's flippers is to propel the animal through the water, they also *used* by the creature for digging holes in which to lay eggs. Millikan tries to accommodate the idea that this digging function of the flippers is also a proper function. The full definition of proper function then reads as follows:

Being built by natural selection is sufficient for proper function, being maintained by natural selection is independently sufficient [like the functional attributes of Swamp Adam and Eve's descendants], and having been *utilized* by other structures built or maintained by natural selection is also independently sufficient for proper function. [My emphasis.] (Millikan (1993) p.49)

Now, the last of these allows novel items like the chameleon's skin colour and the turtle's flippers *qua* diggers to have a proper function but it also follows that this kind of modification does not allow one to discriminate against the Swampman as teleosemanticists wish. For, if I, as a being with an evolutionary history, utilise Swampman, then Swampman has proper functions. If I find SwampDavidson wandering alone and I engage him in conversation, utilising his noises as answers to my questions, then those noises have the proper function of being answers to my questions and so it seems natural to say Swampman has the proper function of

thinking. Something has gone badly wrong with Millikan's account, if I am right in thinking it ultimately leads to one being able to construe Swampman as a thinker.

One possible line of defence that might appeal to some would be to employ some notion of original intentionality here. My relation to Swampman here is similar to that many see between humans and computers. I think any such notion of original intentionality sits very uneasily with the externalist thrust of teleosemanticist thought. No-one can be truly said to have original intentionality: it is all derived from somewhere; either one's ancestors or from someone with ancestors.

Another response one might be tempted to give is that the above quote represents an unfortunate concession on Millikan's part. That is to say, the teleosemanticist should stick solely to the idea of production and ignore cases which seem to involve utilisation. But there are reasons to doubt the efficacy of such a response for teleosemanticists. It does not appear there is any sharp distinction between utilisation and production relations, particularly once we move into the realm of human artefacts and language. As Beth Preston (1998), for instance, points out, language tokens, as events, are *productions* by individual language users, but are also utilisations of pre-existing language devices. Similarly, with Swampman: one can think of his noises as being utilised by me as answers or (at least to some extent) produced by my questions.¹⁷ Or consider a Swampbaby again. Couldn't it be said that after years of training such a baby I *produce* an English speaker and thus a representer of propositional attitudes?

¹⁷ Perhaps in that case we might say then that only Swampman's utterances have proper functions, not the individual.

There is no neat partition of production and utilisation relations which could make this kind of reply persuasive.

Given Millikan's definition, Swampman, I maintain, comes out a thinker. This is evidence of some kind of flaw in Millikan's theory and once we look at her account in detail it is not difficult to see the root of the problem. Millikan's definition of derived proper function seems to be at odds with her general view of proper function. The original account of selection involved simply an appeal to natural selection. The derived notion extended this to a kind of second level selection, selection by an item that arose through a process of natural selection. Millikan introduces this extended notion of proper function in order to accommodate novel functions – something which is essential if mental states are to be considered proper functions.¹⁸ But, and here is the core of the matter, the normative element of the proper function is then lost since there is no history to compare the performance of the item now with the performance of its ancestors. Hence, one of the main motivations of Millikan's account is undermined.

Examples of novel usage stretch far further than thought and language; the same type of explanation is needed to account for both the invention of and improvised use of artefacts. Millikan is explicit on the first of these:

[A] thing that bears no resemblance to a can-opener previously on Earth... may still *be* a can opener [i.e. have that as its proper function], and maybe one despite the fact that it doesn't work. (1993. p.22)

Similarly, since the utilisation of an item can confer a proper function upon the item, it must be true that if I find a stick as I wander through the swamp and utilise it as a walking cane to support my tired body, it does indeed have supporting my tired frame as its proper function. But this obviously has the consequence that anything one *treats* as having a specific function has that as its proper function. Thus by treating Swampman like an *artefact* we ensure that he has all the proper functions we could desire. But then, we remove the idea that what counts as proper or improper functioning is given by the selection history; rather, it is the way we treat objects that confers the particular function upon them. So there is in fact no reduction here at all, the explanation of what does and does not count as a proper use of language or a right representation is given by other, non-biological factors; factors we choose to pick out. That is why, contrary to the expectations of Millikan, Swampman comes out a thinker.

Millikan (1999) has responded to analogous criticisms of her account of proper function. There she maintains that the bifurcated interpretation of proper-function I have outlined above is a misreading. In reality, novel proper functioning items are not a new and special category of proper function but a particular way to describe the same old proper functions.

The job of the pigment arrangers is to produce the relational-structure, skin-color-matching-its-background. But neither *relatum*, brown background or brown skin, is an operative part of an historically normal set of sufficient conditions explaining the capacity of the chameleon to become camouflaged.... Only the whole relational

¹⁸ Christopher Peacocke's (1993) objection that Millikan's account cannot deal with the content of

structure has ... [a proper] function... However, given that brown is the colour of the background, the job of the pigment arrangers is certainly to make the skin brown. (1999, pp.202-203).

So in the case of content bearing states such as beliefs, they can be said to have a proper function because they are particular concrete instantiations of more abstract belief-fixing mechanisms. Crudely speaking there is a belief-fixing something or other (in the brain, say) the proper function of which is to correlate beliefs in the head with environments outside the head. That is its relational proper function as Millikan would say. The belief is true when the mechanism functions normally and false if it malfunctions. Any particular belief is a concrete instantiation of that general belief-making proper function. To speak of the particular belief as having a proper function is only to speak of it as a particular instance of the general relational proper function. So, there is strictly speaking no new function here. Rather what we are doing when we pick out, say, the chameleon's novel skin colour as a having some function is focussing on some particular part of the proper functional story in some particular well-defined case. Where by a 'well-defined' case, we mean a concrete instance of the functioning of the whole abstract relational proper function. But even if we were to grant that there existed a belief-fixing mechanism along these lines, this does not sound like the kind of explanation of beliefs (not to mention other psychological states) that we hoped for from teleosemantics. Teleosemantics held out the prospect of reducing psychological states to biological ones; all this sort of explanation does is provide an explanation of some very abstracted capacities that allow the

accidental generalisations is answered in the same way by Millkan (1993b).

formulation of psychological states. That seems worlds away from providing a reduction of psychological states to biological functions. Consider again Einstein's novel beliefs about space and time. On Millikan's account as presented in these papers, we are to understand this belief as a particular instance of the belief-fixing mechanism operating normally. Such an explanation leaves the most interesting part of Einstein's belief out, the part that we want a reductionist account of content to explain. It does not tell us anything about what is special or unique about Einstein's belief content. As such it can hardly be considered a reduction.¹⁹

We can perhaps make the point more vivid by again contrasting Millikan's account with Kim's. Kim would tell us that there is a set of unique causal inputs and outputs associated with Einstein's beliefs. The analytic functionalist has a problem identifying which inputs and outputs are definitive of the belief. Teleosemantics seems to provide an answer: the selected ones are the relevant ones. But now we see that notion of selection does not apply to individual beliefs and their contents so we are left with at best a proper functional explanation of what beliefs and desires²⁰ do generally. Teleosemantic theory cannot tell the analytic functionalist which inputs and outputs are relevant, it is a story which works at a higher level of abstraction. We don't have the reduction we hoped for.

We must now consider the broader implications these observations have for teleosemantic theories and set this discussion in a sharper ontological context.

¹⁹ This may be considered an idiosyncrasy of Millikan's teleosemantic programme. Other accounts that claim to explain some basic concrete mental states (like Papineau's (1987)) face an analogous problem when it comes to go from these basic beliefs and desires to more complex ones.

²⁰ Or at best with an account of what the proper function of certain kinds of desires or beliefs. See Millikan's (1993) response to Peacocke.

3.2.3 Neander and swampcows

Key to the idea that Swampman does not think is that there is more to semantic content than merely the right physics. Karen Neander (1996) provides a colourful example of how deep this goes for a biofunctionalist. Instead of debating the existence of Swamppersons, Neander discusses the case of a Swampcow – named Craisy (presumably it deserves this name by being molecule by molecule indistinguishable from a British bovine). Craisy, Neander points out, would not be classified as a cow by biologists. All biologists agree that classification of species is, in part, a matter of ancestral history. If one wants to suggest that Swampman is a thinker, then one will also be committed to saying that Swampcow is a cow. (The motivations must be the same for both claims: Craisy looks like a cow, walks like a cow and presumably milks like a cow.) Thus, one will also be committed to saying biological scientists are wrong – high treason for a naturalist. Of course, this argument depends crucially on there being no principled difference between semantic/ psychological functionality and biofunctionality or, in fact, that a functional explanation is the right sort of explanation of psychological states in the first place.

Let me set this thought in a broader metaphysical context. The teleosemanticist claims that biological kinds do not supervene on present physical states. What *constitutes* a biological kind is its evolutionary history. By contrast physical entities, that is entities discussed in the physical sciences, are differentiated by their inner constitution. Even if, as a matter of fact, the explanation of the formation of the said entity is historical, as it would presumably be for the formation of all the heavy elements, that explanation is not part of what

it is to be that entity. The necessity of introducing history is what prevents biology being reduced (at the theoretical, if not the ontological level) to physics.

The teleosemanticist's next move is a reductionist one. They claim that psychological functions can be reduced to biological functions or at least this is their programme.²¹ But if we have good reason to reject this move, if we have reasons comparable to the teleosemanticist's rejection of the reduction of the biological to the physical then the whole programme is undermined. There might be two (not necessarily mutually exclusive) ways in which this is done. Either one demonstrates that psychology is more like the formation of the heavy elements (how the states were formed is not part of that which make them what they are) so there might still be the possibility of a reduction of psychology to physics on such an account. Or that the relation of biology to psychology is a similar kind of non-reductive relation that the teleosemanticist claims exists between biology and physics.

To put it another way, one can agree with Neander that Swampman is not human (and therefore not offend any biologists) but still claim that it is at least an open question whether Swampman thinks, represents or speaks a language.²² One reason both Neander and Millikan offer to reject this move is that psychology is essentially a *human* study and only biology can tell us which species is human. I think this response involves an equivocation. It might well be true that there is some discipline which can be correctly described as human psychology, which

²¹ Millikan is far from clear on this matter it must be said. Sometimes it appears she is suggesting a reduction of psychofunctions to biofunctions, sometimes she appears to be suggesting that the notion of proper function is a highly abstract and general notion which has applications outside biology. Indeed, much of *Language, Thought and Other Biological Categories* can only be understood in terms of this latter interpretation. If that is the case Millikan is merely working by analogy, rather than by reduction.

²² Note, of course, that one can still answer no to this question while disagreeing with the teleosemanticists; that presumably would be the position of the historical externalist – this is why the question is still open.

necessarily involves understanding its subjects as having a certain evolutionary history; modern evolutionary psychology might well be that sort of discipline. However, there may be a more wide ranging science of psychology, one which studies any creature capable of language use or representation, which does not involve deference to any particular biological category. The range of application of biological concepts is not simply given by what is considered physically relevant. Similarly the range of psychological concepts may not be given by what is considered biologically relevant.

Indeed the bulk of the second half of this chapter has been an argument to show the “may” of the last sentence should be an “is”. At the intuitive level, it seems difficult to see why evolutionary history does matter.²³ More importantly, it seems clear that Millikan’s two-level functional analysis cannot accommodate the novel while staying true to its biological reductionist aims. Whatever way the teleosemanticist chooses to cut things up, the theory will founder on the problem of explaining new content. If we are to explain actual concrete instances of new function, then some kind of idea of secondary selection is needed. This undermines the attractive normative aspects of the proper function definition, the very element that made it seem superior to analytic functionalism. That can be avoided by working at a higher level of abstraction, as Millikan has suggested.

²³ This intuition even surfaces in teleosemanticists. William Devries (1996) remarks in a footnote that if Swampmen were common then they should indeed be considered as thinkers. (Papineau (unpublished) has said something similar.) This, admittedly, sounds very strange. Why should the existence or the non-existence of a certain number of other Swampmen affect whether any one Swampman thinks? How many would need to emerge from the swamp before we could safely declare this Swampman could think? I think this kind of intuition shows there is something deeply problematic about the teleosemanticist claim to reduce functional states to biofunctional states. If Swampmen can be thinkers if there are enough them that suggests that all we get from biology is a (reliable) story about the origin of some functional attribute, since that is all we get from the idea that Swampmen are common. This combined with the remarks above and Manning’s (1997) criticisms of proper function suggest there is little to be said for the claims of the teleosemanticist that they are offering a reduction.

Unfortunately, that undermines the idea that teleosemantics provides a reduction of the semantic or intentional aspects of mental states; at best, it appears to be part of a story which explains their presence. This is, of course, not to suggest that there is anything wrong with evolutionary theory. In fact, as I said, I expect there will be insights from psychologists taking an evolutionary approach to our mental life. But they will be insights, I imagine, rooted in an autonomous understanding of psychology. The practice of evolutionary psychologists does not require and could not have a reduction of psychology to biofunctionality.

3.2.4 Why physicalism?

This last chapter has taken us a way from the immediate concerns of physicalism into a discussion of two reductionist projects. One could hold to either Kim-style analytic functionalism or teleosemantic theory and still deny physicalism. For example, one might deny that (MS) correctly characterises the relationship between physics and other non-psychological disciplines while holding Kim's functionalism. Similarly, one might be a teleosemanticist but not a physicalist, if you consider biology to be autonomous of physics.

The failures of analytic functionalism and teleosemantics combined with the discussion of supervenience raise an important question for physicalists – why bother? If even limited reductionist programmes are clearly inadequate and non-reductive accounts lead to contortions regarding the efficacy of the non-physical, one might wonder what the remaining attractions of physicalism are supposed to be. It appears to be undermined by both the messy state of current science and the failure of the metaphysician to accommodate this messy picture in any sort of plausible framework. My best guess is that people think it is forced upon them by

physics; the completeness of physics makes any other position just plain crazy. It is to that thesis, the last pillar of physicalism, that I turn to next.

Chapter 4

The Completeness of Physics

The completeness of physics (CP) is the foundation upon which physicalist metaphysics is built. David Papineau (1993) defines it thus:

I take it that physics, unlike other special sciences is complete in the sense that all physical events are determined, or have their chances determined, by prior *physical* events according to *physical laws*. In other words, we never need to look beyond the realm of the physical in order to identify a set of antecedents which fixes the chance of any subsequent occurrence. A purely physical specification plus physical laws will always suffice to tell us what happened, in so far as that can be foretold at all. (p.16)¹

(CP) promises to provide a quick way to argue for physicalism. As Gene Witmer (2000) has confessed “piecemeal” physicalism – the attempt to show that particular domains reduce to (or supervene on) the physical – looks a poor way to argue. I agree. I hope that what has gone before has shown that evidence does not support this kind of physicalist argument. Nevertheless Witmer thinks physicalists should not lose heart, the quick route to their favoured conclusion is still open – the overdetermination argument. It may be paraphrased as follows:

1. Mental events cause physical events. (For example beliefs and desires cause actions.)

2. Physics is causally complete. (Physical events have only physical causes.)
3. Most events are not overdetermined.

Therefore mental events must be (or be dependent upon or supervene on) physical events²

I think, as it stands, this is a pretty unconvincing argument. There is an obvious tension between the claim that physics is complete and the idea that some non-physical domain (e.g. the mental) causes physical events. In the following I will consider various ways one might define physics so as to make the above argument cogent – I shall argue none succeeds. There is no short cut to the physicalists' conclusion, its plausibility stands or falls with particular claims about the particular relations between physics and other disciplines. However, before treading down that path, it is worth considering some problems indeterminism creates for (CP).

4.1 Causal compatibilism – no chance

The overdetermination argument is standardly presented as forcing a choice between an identification of the mental and physical or massive causal overdetermination³; and the latter is thought to be so absurd as to not require further discussion.⁴ It should, though, be immediately apparent that since there is no determination at all in an indeterministic world, this kind of argument can have

¹ See also Barry Loewer, "Strong supervenience" in Savellos and Yalçin (1995) for a definition along similar lines.

² See also Peacocke (1979) and Jackson (1996). Clearly the argument is a variation on Davidson's anomalous monism.

³ I have some empiricist friends who find this kind of talk of overdetermination embarrassing. They say that physics proper does not make use of the notion of cause and thus all talk of things like overdetermination makes no sense at all. Of course, if you are a physicalist and you don't believe that there are such things as physical causes, it will follow *a fortiori* that the mental is epiphenomeal (that there are no mental causes).

little force. The indeterministic equivalent of determination, namely that antecedent physical events are sufficient to fix the chance of the effect does not readily substitute into the above premises to provide a convincing argument. What, for example, would it mean to say that the chance of any one event was overdetermined?

The question threatens the cogency of the overdetermination argument. Consider Witmer's own analysis of the argument:

C *Completeness* Every physical effect has a sufficient physical cause.

I *Impact* Every seeming non-physical event is a cause of some physical event.

O *Overdetermination* There are very few, if any, events that are causally overdetermined. (p.276)

Now in an indeterministic world the first premise cannot be true; some causes are not sufficient for their effects. The standard picture of non-deterministic causation that is employed is one in which the cause is in some way taken to increase the chance of the putative effect. Hence the general picture one has is of a number of factors contributing to making the event, the putative effect, more or less likely. In this context, it is difficult to draw any parallel with the supposed absurdity of overdetermination. Given that no cause is sufficient for the occurrence of the effect, there seems to be room for other factors to have a real effect on the likelihood of the outcome. So it could be true that all events have physical causes but they might also have non-physical causes as well, without the faintest suggestion of overdetermination. Indeed, for many people (for example Elizabeth

⁴ Although see Crane (1995) and Mellor (1995) for a couple of exceptions to the rule.

Anscombe) because indeterminism seems to leave space for other non-physical, operative causes that contribute to but do not determine the effect, it is an attractive ontological alternative. Of course, it is true that such a model would violate the completeness of physics; even if all events did have physical causes, it would not be the case that the chance of any event was fixed only by physical antecedents. However, it is less clear given an indeterministic universe why this should be thought to be absurd. In short, since there is no determination at all in an indeterministic world it is difficult to see what force the *overdetermination* argument is supposed to have.

I want now, though, to look at another aspect of the completeness of physics in an indeterministic context. It might be argued that the definition of (CP) could solve some of the problems physicalists encountered in chapter 2 – problems with mental causation. A particularly vivid expression of those problems is provided by Tim Crane (1995). Physicalism, he points out, is supposed to solve the problems with mental causation that have plagued dualists since Descartes. Mental causes are just physical causes according to physicalists. However, that argument has been itself undermined by non-reductive physicalism (what Crane calls orthodox physicalism). Since the relation between physical states and mental states is no longer taken to be identity but something weaker, physicalists need to invoke a special and qualitatively different notion of causation to explain how mental events cause anything; and thus the original motivation for physicalism is undermined. In chapter 2, I employed essentially the same line of thought: namely that non-reductive physicalists protestations that mental properties are causally efficacious or relevant or whatever because they fall under patterns of laws or counterfactuals makes them seem causally secondary in some way. That is

to say, the real work of causation is done by the physical events and ‘mental’ causation is a happy accident which falls out of the way we name or locate mental properties in various patterns. It is a mere consequence of genuine causation, rather than causation itself. Such is the basis of the charge of epiphenomenalism.

One worry one might have about this argument is that if physical causation itself can be defined in terms of patterns or counterfactuals, then there seems little ground for insisting that it is substantially, that is to say metaphysically, different from non-physical causation. I shall argue that such a defence cannot work. Given the standard definitions of the completeness of physics and the kind of work it is supposed to do, attempts to give a satisfactory account of non-deterministic causation in terms of probabilities, whether conditional or counterfactual, face serious problems. To make the definition of the completeness of physics workable, in the indeterministic case at least, causation cannot be defined in terms of patterns or counterfactuals.

Let us consider how we might define (CP) in an indeterministic world. What would it mean to say that physics ‘determines the chances’ of all physical events? Perhaps the most natural interpretation of Papineau’s claim would be that physics *fixes the probability* of all physical occurrences. Elliott Sober (1999) provides an explicit account along these lines. According to him the completeness of physics should obey the following relation:

$$(CP) \Pr(B/P) = \Pr(B/P \& M)$$

That is, the chance at time t that B will occur at time $t+dt$ is fixed by the physical probabilities at time t ; the value is unaffected by taking account of the system’s mental properties at time t as well. (CP) says

that the physical properties instantiated at time t “screen off” the mental properties instantiated at that time from behaviours that occur afterwards. (1999, p.4)

Before assessing Sober’s proposal in any detail, a couple of remarks are necessary regarding the interpretation of the probability calculus. Since this discussion is taking place under the assumption of indeterminism, the probabilities being employed must in some way be objective. Furthermore, I think it should be immediately apparent that a frequency interpretation of the probability calculus will not do. Since the frequency interpretation makes the correlation between initial conditions and outcomes depend solely upon the ratio of positive occurrences to actual trials, it cannot do justice to the modal force of the completeness of physics. The physicalist needs to be able to say that it is more than a matter of contingent fact that the physical screens off the mental; a frequency interpretation cannot give her that assurance.⁵

Some other account of objective probabilities is required, if Sober’s definition is to work. Exactly what that would be – some kind of propensity interpretation or something else – does not concern me much here. Whatever your favourite objectivist interpretation of the probability calculus happens to be, it is possible to construct a counterexample to Sober’s proposed definition of (CP).

Consider a situation in which the physical factors X , Y and Z determine the probability of some event, B , occurring. That is to say, given the laws of physics it is always the case that $\Pr(B/XYZ) = 0.8$. It is also true that XYZ “screens off” B from any mental state, M . In other words, $\Pr(B/XYZ) = \Pr(B/XYZ\&M)$.

However, this does not guarantee that mental states are not directly implicated in the causing of B. We might imagine that sometimes when the events X, Y, Z occur they are preceded by another physical event, D. D in turn causes two events, a mental event, M, and a physical event, P. Event M has a positive causal impact on the outcome. That is by itself it would make the occurrence of B more likely. P has a negative impact on the overall outcome of B. It just so happens that M and P never occur in this world without being preceded by D. It further happens to be the case that the amount which M increases the probability is exactly cancelled out by the amount which P decreases the probability. So the probability of B given X, Y, and Z is unaffected by the additional physical factor D. In other words, D and indeed P too are screened off from B. Should we conclude, though, that these three factors, X, Y and Z, are the only relevant causes? I think not, since the way I constructed the situation ensured M had a positive effect on the outcome of event B.⁶

In fact what is going on here is that I am making implicit use of the following *counterfactual* probabilities:

1. $XYX \ \& \ M \ \& \ \neg P \ \Box \rightarrow \Pr(B)$, where $\Pr(B) > 0.8$
2. $XYZ \ \& \ \neg M \ \& \ P \ \Box \rightarrow \Pr(B)$, where $\Pr(B) < 0.8$

In this world, $\Pr(B/P) = \Pr(B/P \ \& \ M)$ holds because $\Pr(B/XYZ \ \& \ M \ \& \ \neg P)$ and $\Pr(B/XYZ \ \& \ \neg M \ \& \ P)$ are not well defined. The probability of M occurring without P is zero; this is what ensures the mental events will be screened off from

⁵ Sober in fact wants to define all the standard physicalist relations in terms of probabilities, supervenience included. He admits himself that a frequency interpretation would fail to capture the requisite modal force of such notions.

⁶ John Dupré (1993, p. 207-8) suggests a concrete example of this. We might imagine that there is a single gene which is correlated both with the desire to smoke and with the desire to exercise.

the physical events. On the counterfactual analysis, on the other hand, it is evident that a mental event may cause a physical event even though it obeys Sober's equation for (CP). Perhaps then if we cannot satisfactorily define (CP) in terms of conditional probabilities, we can do so in terms of counterfactuals. One way to formulate the relevant alternative notion would be:

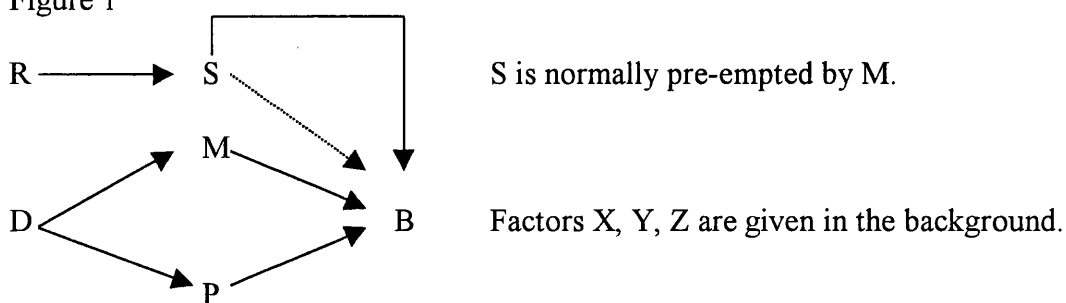
(CP') $P \ \& \ \neg M \ \square \rightarrow P_1(B)$, $P \ \& \ M \ \square \rightarrow P_2(B)$, where $P_1(B) = P_2(B)$. Where B is a physical event and P and M are antecedent physical and mental states, respectively.

In other words, no non-physical factors affect the counterfactual probability of any physical event occurring.

Unfortunately, such a definition will fare no better. Counterfactual accounts of causation suffer well-known difficulties with cases of pre-emption.

Imagine, the same situation as before, where an event D causes two events M and P which raise and lower the probability by the same degree. The above counterfactual relation might nevertheless hold because M might pre-empt another cause, S, which has the same degree of influence.

Figure 1



These two dispositions in turn increase and decrease the risk of cancer by the same amount. But

Hence in the closest possible world to ours, the counterfactual relations hold but intuitively the completeness of physics is violated.

Moreover, it is not clear to me that standard formulations of the counterfactual analysis of causation can avoid the problems of the causal exclusion argument. Lewis's account denies the backtracking counterfactual⁷ in order to avoid the possibility that epiphenomena might be counted as causes. But it seems likely that such a denial will also render supervening events causally irrelevant. That is to say, since supervening events are in causal structure very similar to epiphenomenal events, it seems unlikely that one could rule out one without ruling out the other.

A decision then has to be made for those trying to defend a probabilistic construal of (CP). Either what is and is not probabilistically relevant is given in this-worldly statements of probability or the relevant definition of (CP) involves counterfactually construed probabilistic claims. If it is the former, then one cannot rule out cases described above where physical events 'fix' the probability but mental events are clearly implicated in the causal process. On the other hand, counterfactual accounts of causation are vulnerable to the claim that they cannot pick out the causally relevant features of the world due to problems of pre-emption.

Is it then impossible to say what (CP) comes to given indeterminism? Clearly not, for one is always free to say that whatever the actual or counterfactual probabilities a complete as possible account of the cause can be given by only considering physical events. However, any definition along these lines must admit

would that mean that smoking did not cause cancer?

⁷ If $A \Box \rightarrow B$, does not imply $B \Box \rightarrow A$ according to Lewis

there is *more to physical* causation than is given either in terms of counterfactual dependence or nomological regularities; and thus if physicalists' defence of the causal efficacy of the mental depends on citing counterfactuals or "laws" involving mental properties, they must also admit that if the mental causes anything, then it does not do so in the same way as the physical. In other words, the argument for the qualitatively distinct nature of non-physical causation goes through and an argument in favour of physicalism (or monism more generally) falls flat; namely, that unlike dualist alternatives, physicalism can make sense of mind-body interaction.

Of course, physicalists are always optimistic. There is a lot of work being done on more sophisticated forms of counterfactual and probabilistic causation. (See in particular Noordhorf (2000) and Lewis (2000) and equally see the recent criticism, Ramachandran (2000) and the other papers in *Journal of Philosophy*, April 2000.) Maybe, just maybe, there will be an account which gets everything just right, that defines away causation in terms of some sort of counterfactual or nomological relation. But until that day it seems to me that physicalists have yet to come to terms with the challenges of an indeterministic world; and yet to appreciate the effect it has on some of their favoured arguments for physicalism.

4.2 What is physics? Some preliminary remarks

There is a well-rehearsed argument against (CP) which goes along the following lines.⁸ If by the completeness of physics we mean current physics then the doctrine is almost certainly false. Physics, as it stands now is not complete, that after all is why it is still an interesting area of research. On the other hand, if what

⁸ See Hempel (1980) and Crane and Mellor (1990)

is meant by physics is simply the discipline that is complete or provides a complete account of everything, then (CP) will be vacuous. Either way there is no interesting sense in which we can truly say that physics is complete. I think the argument is broadly right. It is nevertheless worth considering in some detail the various ways in which physicalists have tried to avoid this conclusion in order to provide some flesh to the bare bones I have sketched.

What is needed by an advocate of physicalism is a definition of physics that makes (CP) at once an interesting and substantial thesis and yet not obviously false. There are essentially two ways to go here, each of which either flirts with the Scylla of obvious falsehood or the Charbydis of vacuity. Either one defines physics with some regard to present physics, maybe with some extra spin to avoid the apparent problems of incompleteness or one tries to identify some means independent of current physics to pick out the bases for the physicalist programme. In the first camp we may place philosophers such as Geoffrey Hellman, Andrew Melnyk, John Post and Elliott Sober; in the second David Papineau and Jeffery Poland. I shall argue that their attempts to right (CP) fail. Whenever the notion of the completeness of physics is put to any metaphysical work, as in the overdetermination argument, then the old troubles reappear. Either the position put forward is vague and inadequate to bear the weight of physicalist metaphysics or the claim that physics is complete seems at best unwarranted and at worst false. Either way it undermines the force of the overdetermination argument.

4.3 Quantum quandaries

The first strategy, that of defining physics with regard to what present day physicists practice, is more in line with the naturalistic pretensions of current physicalism. A tempting way for such a naturalistically minded philosopher to finesse the problem of the obvious falsehood of current physics might be to draw on the work of contemporary scientific realism. The scientific realist holds that the theories we have now are approximately true; that is to say, whatever the future development of science, there will be a way of understanding contemporary science such that it approximates the true theory, whatever that might be. Likewise, the physicalist may claim that although current physics is perhaps incomplete, it approximates a complete theory. However, this does not appear to be much of an advance unless we have some idea of the particular constraints this is supposed to place on future physics or some better understanding of what is meant here by approximately true. And as any philosopher of science will admit there is no halfway decent theory of approximate truth.

Really the only way to get this position off the ground is to consider the plausibility of some *actual* physical theory as a basis for physicalism. Physicalists must hope they can motivate the claim that this theory is complete, whatever the state of the rest of physics. The obvious candidate is of course quantum theory since, general relativity apart, it has the best claim to be fundamental; and unlike general relativity it keys into the mereological intuitions of most physicalists.

To assess the claims of quantum theory to ground the physicalist programme will require some explication (but not too technical) of the formalism. Consider an electron. According to physics, electrons have certain properties: energy, mass, spin, momentum, etc. Let us focus on one of those properties, spin. I choose spin not because there is anything physically special about this property

in terms of its quantum description but because it will make the explanation easier. Spin's appeal is that, for an electron at least, it can only assume one of two values (eigenvalues) along any particular axis of measurement: it is either spin up or spin down. The obvious disadvantage of selecting this property is that it has no classical analogue, so it is perhaps difficult to obtain an intuitive grasp of what is being discussed. If the reader finds that he is uncomfortable with the strange property spin, then he may replace talk of spin with one of the other more familiar properties such as position or momentum common to both classical and quantum physics. Nevertheless, he should appreciate that although this will not change the nature of my analysis, it would make the examples more complicated since, in the case of the other properties, the electron has a large or in some cases infinite number of possible values it might take on.

In order to discern the value of spin for a particular electron, we must choose some axis of measurement. By convention we name one such axis the x -axis. Let us imagine we have just measured the spin of the particle along this axis and we have discovered the electron is in a spin up state. We label the quantum state of the electron which corresponds to that result $|Up_x\rangle$ (and we call this an "eigenvector"). Quantum measurements are such that if one were to remeasure immediately the spin of the electron along the x -axis one would always obtain the result, spin UP_x ⁹, since the particle remains in the state $|UP_x\rangle$. No great surprise. However, if we now take the same electron and try to measure the spin of that particle along some different axis, say the y -axis, things start to become stranger. Sometimes when we measure the electron it will be spin up, that is in state $|UP_y\rangle$, and sometimes it will be spin down, that is in state $|DOWN_y\rangle$. A natural

conclusion to draw from these results would be that some electrons that have the property UP_x also have the property UP_y and some have the property DOWN_y. One would then expect that, if we measured the electron along the x-axis of spin immediately after it had been measured along the y-axis, we would always obtain the result spin UP_x. This is not the case. The measurement of the system along the y-axis in some way disturbs the quantum system so that the outcome of measurement along the x-axis becomes totally randomised, sometimes one obtains the result UP_x and sometimes the result DOWN_x. This strange effect is a consequence of what is known in quantum theory as the *incompatibility* of the two variables, spin along the x-axis, and spin along the y-axis. Other pairs of properties are also incompatible; for example position and momentum, the various components of angular momentum with each other. It is impossible to obtain definite values for any such incompatible pairs simultaneously. So given that we have obtained a definite value for spin along the y-axis, we cannot provide a definite value for the spin along the x-axis, it is neither spin UP_x, nor spin DOWN_x. However, one can represent the state of the system for the spin x-component after such a measurement, in the following way: $A(a|UP_x \rangle + b|DOWN_x \rangle)$, where, a and b are constants and A is some normalisation factor. This is called a superposition of states. Although the superposition itself does not represent a definite value, it does allow one to predict precisely the probability that either an $|UP_x \rangle$ or $|DOWN_x \rangle$ state will occur after measurement. It is these superposed states that I think are likely to cause some initial problems for physicalists.

⁹ I hope the nomenclature here is clear: the unbracketed state represents the value after measurement, the bracketed state is the wave equation.

A first naïve argument against physicalism might be to say that superposed quantum states do not represent definite properties and hence cannot form the bases for a physical description of the world. As Jeffrey Poland puts it in defining his physicalism: “[I]f there is no fact of the matter regarding as to what objects are included in the bases, then there is no fact of the matter regarding what the theses [physicalism] are expressing, and hence there is no fact of the matter regarding whether the theses are true or false. Vacuous or indeterminate content, therefore, undermines the significance of physicalist doctrine and obstructs the attainment of the goals of the physicalist programme.” (1994, p. 148)

The anti-physicalist might argue then that because superposed states are value indefinite, there is no fact of the matter concerning whether the electron is in one state or the other. Hence, a physicalist could not assert, say $P = m$ or P supervenes on m , where P is a physical property and m a non-physical property, because P is not well defined for any superposed states (and all quantum states are superpositions under some description). But Poland has an obvious reply: “the wave function gives very definite descriptions of the system in terms of the probability distributions; such descriptions give a complete characterisation of the system. Thus for any such description or associated probability distribution there is a fact of the matter whether it is in the physical bases. What more could anyone ask for in characterising such bases?” (1994, p. 174)

In other words, Poland is prepared to characterise the physical bases as *completely and determinately* described by the superposed states.¹⁰ Now, this may answer the naïve objection that the bases are ill-defined but it simply gives rise to a more profound problem. Namely, if superposed quantum systems are the base

physical states, then how do they give rise to macroscopic phenomena which never appear to be in superpositions but always in some definite property state? We seem to have simply reversed the direction of the mystery. Looking down from the macroscopic domain to the quantum domain, one cannot understand how quantum states can be well defined. Looking up from the quantum domain, we cannot see how to recover the macrophysical image.

Scott Sturgeon (1998) has made essentially the same point in the context of the overdetermination argument. He argues that inasmuch as we think it is at all plausible to consider physics complete, we generally have in mind high-level theories with well-defined dynamics like quantum theory. That is to say, we do not imagine that physicists will need to invoke non-physical causes or explanations when they do their experiments in the laboratory. However the sort of physical events that we think are caused by mental events are macroscopic physical events. And since it is not at all clear how the macroscopic and the microscopic domains relate, because of the conceptual strangeness of quantum mechanics, the overdetermination argument is no good.

Indeed, as Sturgeon suggests, raising the question of how the micro and macro images relate is a way into what is known as the “measurement problem” in quantum theory. It is perhaps easier to appreciate the force of Sturgeon’s objection in terms of that problem. To explain the details, we need a richer account of quantum theory than has been given up till now.

Quantum mechanical states are governed by the Schrödinger equation, which tells us how some state, say, $|UP_x\rangle$, develops through time. After interaction with the environment $|UP_x\rangle$ will develop into some superposition of

¹⁰ Poland also suggests that the definite value states acquired after measurement could also form

$|\text{UP}_x\rangle$ and $|\text{DOWN}_x\rangle$ states. However, on measurement, as was implicit in my earlier discussion, the wave equation “collapses” from a superposed state to an eigenvector either $|\text{UP}_x\rangle$ or $|\text{DOWN}_x\rangle$. The problem here is that the measurement interaction is not part of the Schrödinger evolution of the system; it is a new type of interaction. If one treats the measuring apparatus like a quantum system that interacts with the electron, then one does not obtain a determinate property state but an entangled system of the form:

$|\text{M}(\text{UP}_x)\rangle |\text{UP}_x\rangle + |\text{M}(\text{DOWN}_x)\rangle |\text{DOWN}_x\rangle$, where $|\text{M}(\text{UP}_x)\rangle$ and $|\text{M}(\text{DOWN}_x)\rangle$, represent the states of the measuring apparatus for measurements of $|\text{UP}_x\rangle$ and $|\text{DOWN}_x\rangle$ respectively. But of course we never do observe situations like this. In fact, we probably cannot even begin to comprehend how we could observe superpositions of states. Measurement then appears to induce a different sort of interaction not captured in quantum theory. The conceptual problem is how to understand what is special about measurement interactions that takes particles from superposed states to definite property states. And this is just the problem of how quantum systems relate to macroscopic systems, since the measuring device is a macrosystem.

Some of the solutions mooted both by philosophers and by physicists to account for wave function collapse should be disturbing for a physicalist. For example, Neils Bohr thought that collapse was induced when the system interacted with an irreducibly classical macroscopic object. He thus denied the possibility that one could apply quantum theory to the measuring apparatus. Irreducibly classical features that are always external to quantum theory, according to Bohr, govern measurement interactions. Hence, the completeness of

base properties. However, one should remember they are only definite states from one perspective.

quantum theory is undermined.¹¹ A similar but more extreme view, held by Eugene Wigner¹², attributes collapse to the interaction of the human mind with the system. Again, non-physical, in this case mental, causes are taken to affect quantum events. Again completeness is undermined.

The above views both attribute collapse to some non-quantum element of the measuring situation. Thus both seem inevitably to lead to viewing quantum theory as incomplete. However, there are interpretations that deny this duality between quantum system and measuring system. Such interpretations, inspired by the seminal work of Hugh Everett III, all deny there is any such thing as wave collapse. The great advantage of these interpretations is that they only utilise the dynamics of the Schrödinger equation. On the other hand the major problem for such accounts is to explain why we think we see collapses and never superpositions.

Explanations have split into two competing accounts: the many worlds view and the many minds interpretation. The former interpretation states that during measurement interactions the universe splits into many universes, each one corresponding to each possible outcome of measurement.¹³ The latter formulation makes the comparatively speaking more modest suggestion that the human mind splits when measurement takes place.¹⁴ So, although on both accounts the universe is actually in a superposed state because my mind or body splits during measurement, I am confined to represent the world in a nonsuperposed way. There are technical problems with both approaches; in particular the many worlds

For an incompatible variable such terms will be superpositions of states.

¹¹ Note, I mean completeness here in the sense formulated by Papineau in his description of the completeness of physics, *not* the sort of completeness of quantum theory that is discussed in EPR-type phenomenon.

¹² See Hughes (1989) for a discussion of Wigner's and Bohr's position.

¹³ DeWitt, B. (1970)

interpretation struggles to make sense of the probability assignments in quantum theory. But such technical problems are not, I think, the cause of the resistance amongst philosophers of physics to either view. I take it that what most find objectionable, and no doubt what any lay physicist will think too, is that such views are fantastical. Who could believe in splitting worlds or minds? But what that strong intuitive objection expresses, I think, is a dissatisfaction with these theories' attempts to explain the emergence of the macroscopic image. In fact, both interpretations make our perception of the macroscopic world false (or at best seriously incomplete). They are, in other words, eliminative about the macroscopic image of the world.¹⁵ I doubt it is the sort of world-view most physicalists will find satisfactory. Furthermore, it is doubtful whether such theories have really explained measurement interaction and the connection between the quantum and the macroscopic level. What for example is so special about measurement interactions that they cause splits in the universe or minds. Why do splits not equally occur at lower levels of reality? In short, the proposed solutions to the measurement problem seem either to lead to the imposition of non-quantum causes or to understanding the world as completely different from the way we think it is. But neither type of view elucidates the relation between the macroscopic and the quantum. One simply asserts there is a strong dichotomy and the other fails to give an adequate answer to why we think there is a strong dichotomy.

¹⁴ See Albert, D (1992). Chs. 6 and 8

¹⁵ In fact, the many minds view is consistent with the universe being in a vacuum state, that is empty although because of the superposed states we think otherwise. Of course, this raises the question what a mind is supposed to be in the many minds interpretation and if it could be accommodated into a broadly physicalist ontology.

To leave the discussion here, as Sturgeon does¹⁶, would provide an incomplete picture. There are other more radical theories of quantum phenomena that do not create the problems I have outlined above and that should be discussed. The Bohm theory is perhaps the most heterodox. Bohmian mechanics posits an ontology of particles with definite position and, like standard views, a wave function. Again like standard views, the wave function evolves according to the Schrödinger equation. The position of the particle is governed by what is known as the ‘guidance equation’ which is dependent on the value of the wave function. The particles always have well defined positions and trajectories, governed by an entirely deterministic dynamics; probability only enters because of our ignorance of the initial states of the particles. (The theory is cooked up in such a way to make sure this kind of ignorance is unavoidable, so there are no empirical disparities between Bohm theory and more orthodox interpretations.) There is no collapse of the wave function, it continues forever to evolve in line with Schrodinger’s equation. Nevertheless, there *appears* to be a collapse because the wave equation guides the particles along a particular path, which depends entirely on the particles’ initial state and this is what is registered when measurement takes place. Strictly speaking all measurements are therefore measurements of position. Spin is not a property of the particle at all; rather it is a way of describing the wave function which guides the particle along certain trajectories when certain types of measurements are made. Hence all properties apart from position are, as they say in the jargon, ‘contextual’; that is partly a consequence of the way a measurement is made.¹⁷

¹⁶ Indeed as I also did in my M. Phil (1998, University of London) thesis, *The Limits of Reductionism*, ch. 3.

¹⁷ A good clear, elementary discussion can be found in Albert (1992), ch.7

Because all the particles have well-defined positions it is easier to see how the macro and micro realms relate. Bohm theory then looks a more promising alternative for physicalist metaphysics. Appearances, though, are deceptive. Although measurement interaction appears less curious, and therefore the connection between the micro and the macro worlds seems less problematic, in Bohmian mechanics, there are grave difficulties in extending the approach to quantum field theory. In essence it is practically impossible to come up with anything that will take the place of position in the context of a field theory. So when we reach deeper into the physics the classical picture breaks down again. Even if the Bohm theory is conceptually less puzzling than standard quantum mechanics, it is not explanatorily adequate. Moreover, even if we were to grant that this were the correct interpretation, it does no more than orthodox quantum mechanics to explain how it is that any of the features distinctive of macroscopic phenomena, from chemical bonds all the way up to people, thoughts, and societies emerge. This is particularly problematic for Bohmians given the contextual nature of all properties other than position.

What the case of the Bohm theory does show, though, is that it is wrong to think of quantum mechanics as creating *special, distinct* problems for the physicalist as perhaps Sturgeon's treatment would suggest.¹⁸ Rather what we have here is just a particularly vivid case of a general problem for the physicalist. Once we fix on some particular theory as our base, then consideration of current science

¹⁸ Sturgeon argues that we cannot be certain that quantum mechanical events compose macroscopic macrophysical events because it is not clear that the quantum mechanical parts are essential to the macrophysical wholes. Here, he draws from the work of Yablo (discussed in the previous chapter) the idea that an event, A, is only the cause of another, B, if A brings about B's essence. As I suggested earlier I am quite sympathetic to Yablo's treatment but I think one may circumvent this talk of essences. What both Yablo's and Sturgeon's work suggests is the much more mundane thought that we do not consider one event to be the cause (or part of the cause) of another unless we have some story, some worked out theoretical treatment which can make sense

makes it highly implausible that the theory in question is complete in the sense required: different, unreduced theories must be invoked to account for a whole host of phenomena.

What *is* special about certain *interpretations* of quantum theory is the idea that the theory itself is incomplete is implicit in our understanding of it; and this in itself is instructive. One metatheoretical reason some physicalists suggest that physicalism is the best metaphysical theory is that it promotes the idea that one should look for underlying mechanisms to explain phenomena. As I stated in the opening sections, I do not think this is a very good argument. One does not need the metaphysics of physicalism in order to promote the idea that looking for underlying causes can often be a successful methodological practice. Moreover, it is certainly far from obvious that such a strategy applied generally in science would provide any backing for physicalism. The underlying mechanisms that scientists investigate are as theoretically diverse as the phenomena they are investigating. What certain views of quantum theory certainly do show us is that one can conduct perfectly good science without such commitments. It is arguably the lack of such metaphysical presuppositions that enabled scientists to develop quantum field theory. (See Fine (1986) for just such an argument.)

A second important feature of a better understanding of quantum theory is that it provides an antidote to what one might call naïve or vulgar materialism. Despite the much-vaunted naturalistic pretensions of present day physicalists, many, I believe, still conceive of the world like the old atoms-in-the-void materialists. In particular, composition or reduction is, in principle, taken to be no more problematic than the building of a wall from bricks.

of that claim; and we don't have that for the relation of quantum events to macroscopic events. This is just the less metaphysical point I'm making.

Quantum mechanics should dispel such complacency. Indeed our discussion of quantum mechanics provides us with an opportunity to highlight further some of the problems in relating different domains that I discussed in chapter 1. Recall that I pointed out that quantum chemistry makes ineliminable use of the structure of the molecules that it is investigating. Our discussion of quantum mechanics can give these observations greater depth.

For the modelling of many molecules a pure quantum treatment would be useless, even if possible. Chemists are interested in the shape of molecules. In particular, certain molecules can take on different shapes: C_6H_6 , for example can take several different forms, including the classic benzene ring structure. As I suggested it is this kind of knowledge that is the basis of quantum chemical calculations. However, the very idea of structure or of the same constituents arranged in different structures is impossible to recover from a purely quantum description of the situation. At best, one would be able to write down some highly complex superpositions of states for any given molecule but you would have no way of determining which particular structure or shape the molecule was in. Given chemists are often interested in investigating the particular properties of different chemical structures such superpositions would be as useless as the superpositions involving quantum system and measuring apparatus are for any real scientific investigation. The quantum treatment could tell them nothing about the *particular* structure they were investigating.

The lesson is straightforward enough: the relation between the quantum domain and other domains should undermine the simple-minded mereological intuitions that underpin much physicalist thinking. If physicalists reject naïve physicalism, then microphysics should offer them little comfort and much

disquiet. Quantum theory provides a clear demonstration of the implausibility of physicalism when confronted with real science.

4.4 Physicalism without physics

How about the alternative ways of defining physicalism? First let's consider David Papineau's suggestion that we try to say what is *not* physical and define (CP) in those terms. In his book *Philosophical Naturalism* (1993) he suggests that whatever the physical is, it is not the mental. The overdetermination argument then runs in the standard way. It is clear that mental events cause physical events; hence barring overdetermination the mental must be (or be dependent upon) the physical. As Elliott Sober (1999) has suggested it is not clear that this is much of an advance as it seems equally unclear what the boundaries of psychology are. Some of the arguments I presented in chapter 3 against functional reduction of content would lend support to that view. Actions, I argued, are individuated in terms of their motivations and relation to external objects – do such actions count as physical or psychological on Papineau's account? I shan't pursue this particular point at any length; I wish simply to reiterate the claim which I raised at the beginning of the chapter.

Any apparent case of a *non*-physical event causing a putatively physical event must make us suspicious that Papineau-defined physics is actually complete or that the putative physical effect is really physical in the desired sense. To put the same point somewhat less tendentiously, given Papineau's definition of physics and the apparent causing of physical events by non-physical events, there is more than one option before us. We might: a) agree with Papineau that non-physical events are really physical events; b) think that the supposed effect is not

really physical; c) that the supposed cause is not really non-physical; d) that physics is not really complete. Perhaps (b) or (c) seem unreasonable for all putative cases of non-physical events causing a physical event, particularly given Papineau's liberal construal of physics. For example, because I believe that it is raining, I pick up my umbrella. The movement of the umbrella can be construed as a physical event; and thus we have a case of mental events causing physical events. Granting this, though, it is not clear why either (a) or (d) is better (or worse).

What is needed to make Papineau's argument go through is a defence of the claim that everything minus the psychological is complete. Papineau's overdetermination argument, as it stands, is at best a way of stating rather than arguing for physicalism, viz. that there is a domain which is complete and all other domains must depend upon it in some way. But without any real argument for why we should believe that physics so described is complete that's all it is – a statement of his metaphysical position.

Indeed once we think of physics as something other than our best current theories and instead conceive of it as mish-mash of generalisations from all disciplines, including commonsense, it is difficult to see why we should believe that kind of physics (everything minus psychology) *is* complete. After all, there are certain kinds of events that we can only explain by citing psychological causes. Missing out such states from our definition of what counts as physics would seem to render the completeness claim false. For example, if we can only explain why Tony stopped to look in the window because he was thinking of buying a *rolex*, then it is natural to say that it is necessary to cite such a cause, if we want a complete description of the kinds of thing Tony does. It seems arbitrary

to exempt psychology from one's definition of the complete science, unless we demonstrate that it is reducible to or eliminable from the rest of the mish-mash without loss. But I take it that our dependence upon everyday psychological explanations shows we cannot get by without psychology or reduce it to some other discipline. So there seems no motivation for claiming that this kind of physics is complete.

Similar remarks also apply to a more recent short paper co-authored with David Spurrett (1998). Here Papineau claims that a more generalised form of the overdetermination argument can be used to argue for separate metaphysical theses. Instead of worrying specifically about what is meant by physics we may claim instead that various kinds of domain are complete. For example, one might claim that the quantitative domain is causally complete and by plugging that into an overdetermination argument show that the non-quantitative is metaphysically dependent on the quantitative.¹⁹ (The other example they provide is again the idea that the non-mental is causally complete.) Again, though, one should not mistake these for *arguments*. Without any motivation for saying why the domains so described are in fact complete, they are simply statements of various features of a broadly physicalist position.

The only arguments I can divine in this paper are points against dualism and panpsychism (p.27-8). This I find curious. Arguments against dualism and panpsychism are just that – arguments against two other metaphysical positions. Even if sound, they do not establish physicalism. It is true that the anti-physicalist has some kind of explanatory debt which must be paid or written off somehow:

¹⁹ Again, of course this presupposes there is a sharp boundary between the quantitative and the qualitative. Are considerations of structure used in physical chemistry to provide models of the phenomena, qualitative or quantitative? Is the complete science to be purely quantitative? What

namely to answer the question if not physicalism, what else. I shall do my best to address this challenge in the second half of this work. Nevertheless, we still need some positive argument for physicalism.

In yet another paper, Papineau's (1998) reason for concentrating on the deficiencies of dualism becomes more apparent. It is worth quoting the relevant section.

I think in the end [(CP)] must be upheld. The question you need to ask yourself is whether particles of matter ... are ever caused to accelerate by conscious causes alone, in the absence of any other force. In effect, do we need to include purely conscious causes alongside gravity, the electroweak force, and, so on, in the category of fundamental forces?

What appears to be at stake for Papineau when it comes to deciding your metaphysics is the number of fundamental forces you believe there are. Fixing the definition of physicalism is just a matter of fixing the total contribution of all factors into a giant force function (or Hamiltonian). Disputes between various metaphysical positions are just then a dispute about the number of different fundamental forces that exist.

If we think of all causes as Papineau suggests in the quote, then we have two options. Either we say mental causes are identical to physical causes or we have to posit an entirely novel and weird type of mental force. But this way of thinking of causation seems to me a retrograde move for Papineau and his attempt to make the overdetermination argument good. The advantage of the woolly

would this mean? Spurrett and Papineau provide no general characterisation of the quantitative domain.

conception of physics with which he began, was that it was clear that mental events did have same impact on those kinds of physical events. However, when we begin to think of causation in terms of *fundamental forces* that intuition evaporates.

For example, my belief that philosophy is not the kind of thing one can make money out of might cause me to seek advice from a career's counsellor but can that effect be thought of simply in terms of the acceleration of particles of matter? Clearly in some sense it must involve redistributions of matter but at the very least it just begs the question to insist that it can be cashed out in terms of Papineau-style physical forces. Even if we added extra non-physical mental causes into the picture that would not make it any clearer that an explanation in terms of redistributions of matter was capturing *everything* that was going on. When we think of a belief causing an action we do not conceive of it as some extra bump given to the physical factors present.²⁰ The question the physicalist must address is how does our actual idea of mental causation fit into their metaphysical picture. It is at best an unfortunate caricature of anti-physicalism to suggest it implies that there are separately functioning mental forces which need to be taken into account when trying to understand how particles of matter accelerate.

Papineau's argument employs a very narrow conception of what is to count as a cause; one drawn for some idealised version of theoretical physics. As I have said it is not the sort of picture of causation that fits mental causes; and that must undermine our confidence in premise one of the overdetermination argument. Why should we say the mental causes the physical, if the picture of

²⁰ It is possible that Anscombe thinks of it in this way though.

causation we are working with is one of fundamental forces accelerating matter? Worse than that, though, for Papineau his account of causation is not the sort of picture which fits physical causes either. As the discussion of classical and quantum mechanics made clear, the relation between different these domains is not one in which we can think of them making contributions to one giant force function or Hamiltonian. The measurement problem in quantum mechanics has nothing to do with thinking of the classical domain as providing an extra bump to add to the sub-atomic and gravitational forces. The problem in relating these two domains is precisely that there is not some general story to tell that can make sense of both the quantum and classical features of any system at the same time. The lessons of quantum mechanics not only apply to those who want to define physicalism with reference to something like current science. They also provide a warning to any defender of physicalism who imagines that whatever the particular relations between domains are, there is a simple, general story to tell about how we can possibly imagine everything fitting together.

Papineau makes the question of the completeness of physics and thus physicalism turn entirely on how bits of matter (whatever they are) are accelerated. In other words, he puts the question in terms of a particular sort of (vulgar) materialism. Of course, if you think of the matter (so to speak) in these terms then anything other than materialism is going to seem absurd. Self-evidently, if you think of causation only ever on the model of physical causation, then physicalism will seem the only option. But we, as a matter of fact, do not do that nor in reality *can* we do that since there is no single picture of causation that emerges from physics itself as the discussion of quantum mechanics highlights.

The way we think of one event causing another is richly pluralistic, varying from one case to another.

To make the joint claim that physics is causally complete and that the mental causes physical events (where physical here must be understood in the same way as it is in the purported completeness claim) at all plausible, we would need a treatment of the physical which allowed us to recover the psychological. But as chapters 2 & 3 demonstrated, we don't have that. So we are left with our ordinary and diverse causal talk which provides us with no reason to draw the substantial metaphysical conclusions advocates of the overdetermination argument desire.

I now wish to turn to the consider Jeffrey Poland's definition of physics. He takes a more subtle line to avoid explicit commitment to the ontology of current science. The idea here, as far as I understand it, is that although there may be no fixed, determinate physical theory from which we can assess the likely truth or falsehood of the completeness of physics; this fact by itself does not mean that physicalism is without content. Physicalism, so understood, is not a theory about how any particular physical theory stands in relation to other domains, rather it is a view about a more abstract kind of thing called the *research programme of physics*. Whatever physics in the end turns out to be, we have a firm grasp on the sort of questions and problems physics must solve. The completeness of physics is thus the claim that there is a theory which answers those questions and that theory is complete.

Obviously what needs to be filled in here is some idea of what the 'research programme of physics' is and what constraints that places on physics. The dangers are apparent. A Quinean definition of the goals of physics as the

science that aims for full coverage would collapse this definition of (CP) back into an analytic truth. Poland comes up with the following four questions which he thinks at least partially define the research programme of physics.

- *What are the fundamental constituents of all occupants of space-time?*
- *What are the fundamental processes that underlie all causation and all interaction between such occupants?*
- *What parameters are relevant to the unfolding of all space-time and hence to all change?*
- *What is the nature of space-time itself, its origin (if it has one), and its destiny?*

(1994, p.125)

Now, this is clearly an advance on Quine's formulation. It certainly does not follow trivially from the above that whatever deserves the title of physics will be complete; and unlike Papineau's original proposal it clearly provides some rough outline of what is meant by physics. Nevertheless, I don't think Poland succeeds in providing an interesting formulation of physicalism.

The term fundamental is problematic for a start. Consider the case of psychology. Clearly people are occupants of space-time but what are their *fundamental* constituents? Well, that is not a question which can easily be answered in advance of a particular theory about how it is that people interact in space-time. As things stand at the moment the best way to account for the interactions of people is by ascribing beliefs and other mental states to them. Either then these properties are fundamental, in which case what is standardly

regarded as the psychological is part of the physical, or they are not. If they are then this definition of physics again looks vacuous. If they are not, then this simply courts the problems discussed earlier in the chapter. Given what we know about the world and the way we describe it, there seems little reason to think that any description of physics as standardly conceived will in fact underwrite all other interactions in the way Poland maintains. Poland's definition, without an account of what he means by fundamental, seems stuck on the horns of our original dilemma.

The definition is also in some other ways too narrow. The only entity (or sort of entity) that Poland's four questions commit him to is space-time. But that should immediately raise concerns in light of our discussion of quantum theory. In particular with regard to point one, superpositions of states (even position) are not happily thought of as occupants of space-time. Even on the Bohm theory, which arguably does more to satisfy classical intuitions than any other, the wave function, the physical guiding force of all the particles does not exist in space-time but in configuration space. Point three is also problematic. Measurement, at least on standard interpretations, involves a change (from a superposition to an eigenstate) but that is not happily thought of as "part of the unfolding of space-time".²¹ Poland's characterisation then suffers from all the problems we have previously. It remains vacuous because there is no, non-question-begging explication of what is meant by fundamental; and it seems too restrictive as it limits the arena of interaction to space-time and that seems problematic on any understanding of quantum theory.

²¹ Not to mention the problems created by EPR phenomena.

When one considers what one wants (what one should demand) from a definition of physicalism, namely that it be non-vacuous and not false, it is not surprising no satisfactory definition can be found. In as far as the doctrine states anything remotely concrete about what physics means, it is bound to founder on the obvious problem that completeness claims are implausible given the current disordered state of science. What we see in fact is that contra Witmer there is no easy route to physicalism since there is no plausible and interesting way to motivate the completeness of physics given what we know about physics.

4.5 Keeping out the crazies

Some physicalists will no doubt just shrug their shoulders at this thought. “Look,” they might say, “okay there is no *easy* way to argue for physicalism. But you haven’t given us any real reason *not* to believe in physicalism. All we are saying is there is some physical theory, maybe we have it now, maybe it’s a bit different which, if we could work out all the detail properly we could show is complete. All you’ve shown is that we do not *yet* have any theory like that, worked out in that kind of detail. But that does not mean there is not one. After all, what are the alternatives?”

I gladly grant that I have not refuted physicalism. I’m not sure how one *refutes* any metaphysical position. What I have tried to show is that physicalism is *not* naturalistic metaphysics. It is a doctrine that when any content can be pinned down at all seems implausible given the way the *sciences* currently represent the world.²² Of course, physicalists are free to pray for a brighter future, with a well-

²² Even when the content of what physicalism is can be pinned down the doctrine suffers from another kind of vacuity: namely, nothing concrete can be said about what kind of relation exists between the fundamental theory and the higher theory.

defined fundamental physics and clear reductions of all other disciplines. The arguments, though, and the state of science lend their position no support.

What then explains the prevalence of physicalism in Anglo-American philosophy for the last fifty years? An interesting paper by Andrew Melnyk (1997) provides an answer. Melnyk I think would agree with pretty much everything I have said: the only way to give content to physicalism is by defining it in terms of current physics; when we do so the doctrine looks implausible. However, according to Melnyk that does not mean we should not be physicalists. What we must consider first, before rejecting physicalism, are the relevant alternatives that try to discharge the same explanatory task. (Melnyk is a little vague on what precisely the explanatory task is.) If there are no competitors that are more plausible, then it is rational to back physicalism (even if one assigns it a low probability). The details of Melnyk's argument and his particular assessment of how one should assign probabilities to metaphysical theories are not important. What the general nature of this argument highlights is one of the deepest of all urges to be a physicalist, one which we have also seen in Papineau's writing: physicalism has to be right because everything else is just plain crazy. It is against this attitude that the rest of the thesis shall be directed.

Section 2

Pluralism

Introduction to Section 2

An open mind is no substitute for hard work

Nelson Goodman

If not physicalism, what then? There remain the three other metaphysical positions outlined at the beginning of this work: idealism, dualism and neutralism. Here, though, I wish to explore a fourth alternative – pluralism. ‘Pluralism’ is a word with many philosophical and perhaps many not-so-philosophical uses. Construed as a straightforward alternative to physicalism, pluralism involves taking the world to consist of many and varied things – properties, laws, whatever – which escape the kinds of dependency on the physical that physicalists have thought unavoidable. In short, one might call this kind of pluralism, *metaphysical* pluralism. A philosophical exposition of metaphysical pluralism would be expected to achieve two tasks. First, to see if there is any such thing. That is, whether conceptual space will allow room for a metaphysics which is not a form of dualism or one of the monisms. The lack of a late, great philosophical heavyweight to champion the cause may make some sceptical that this task is possible.¹ Even if it is, then the second task is even more daunting: to argue that pluralism is true.

I shall not here undertake either task, in any straightforward sense. Rather, I shall develop what I consider a viable pluralist alternative through the work of four contemporary philosophers: John Dupré, Nancy Cartwright, Nelson Goodman and Hilary Putnam. Dupré and Cartwright offer a scientifically motivated pluralism; Goodman and Putnam a pragmatist version of the same. The

¹ Apologies here to William James and his followers.

following chapter will discuss the works of Dupré and Cartwright, the next the pragmatic pluralism of Goodman and Putnam. The aim will be both critical and constructive. Critical in the sense that both chapters will attempt to point out the weaknesses of each thinker's position. Constructive, because what is positive will be used to form a tenable version of pluralism. As will become apparent, there is much that all four philosophers actually have in common, and the differences in their positions, such as they are, as often as not, concern matters of emphasis and interest, rather than of principle. But before I begin with the detail, a broad distinction introduced by Huw Price (1992) will aid the discussion.

Price distinguishes two types of pluralism – horizontal and vertical. Horizontal pluralism is the view that there are many equally apposite ways to carve up the things of the world. Quine's famous (or perhaps infamous) doctrine of ontological relativity is a version of horizontal pluralism *par excellence*. Vertical pluralism, on the other hand, is the idea that there are many different levels of description of the world, corresponding to many different types of thing, all of which have the same ontological status. The type of pluralism that I am interested in, it should be clear, is the vertical sort. And indeed, I take it this is the sort of pluralism that Dupré (certainly) and Cartwright (possibly) are anxious to defend. What I hope to show through the work of the four philosophers I discuss is that a convincing version of pluralism must incorporate elements of both vertical and horizontal types. Through the work of Putnam, in particular, I hope to show that a version of horizontal pluralism can be made plausible.

In the concluding chapter, through a discussion of Donald Davidson's Anomalous Monism, I shall suggest how to go from horizontal to vertical pluralism. By the end, it should become clear that my position is as much a rejection of anything

that might reasonably be called *metaphysical* pluralism as it is of physicalism. My real objective will be to show that if we reject physicalism and embrace pluralism, in my sense, we are not attempting to provide an alternative *metaphysical* picture but rather rejecting the idea that metaphysics, any metaphysics, can provide the very abstract and general answer to the question, what is the world like.

Chapter 5

Science without Laws

The following chapter will discuss the works of John Dupré and Nancy Cartwright. Both argue to a kind of pluralism from detailed studies of the sciences; biology in the case of Dupré, physics in the case of Cartwright. Both converge on a similar view of causation that is, *prima facie*, at the heart of their pluralistic positions. The emphasis here will be more on what I think is wrong with their respective positions; where they fall short in offering satisfactory explanations of what pluralism is or why one should adopt it. But that should not obscure the fact that I agree with a great deal of what both say about the sciences. Indeed, certain remarks of Cartwright's provide an important insight into how pluralism can be developed and defended. And both represent what to me is an important general attempt to give a 'face-value' reading of the sciences and their relations to one another, unencumbered by traditional metaphysical thinking. In that respect, perhaps more than any other, they fight the good fight. However, as shall become clear in later sections, they lack a certain philosophical, or perhaps even metaphilosophical, attitude which needs to be developed explicitly before pluralism will seem persuasive.

5.1.1 Ontological democracy: John Dupré's pluralism

John Dupré offers up two types of pluralism in his work *The Disorder of Things*, each a consequence of the rejection of another doctrine.

Pluralism, first in opposition to an essentialist doctrine of natural kinds [is] the claim that are many equally legitimate ways of dividing

the world into kinds, a doctrine I refer to as “promiscuous realism”; and second in opposition to reductionism, pluralism [is] the insistence on the equal reality and causal efficacy of objects both large and small. (1996, p.7)

Dupré’s arguments for the former doctrine are based mainly on an analysis of the problems of taxonomy in the life sciences. The latter is defended by various arguments against reductionism, in tandem with what Dupré calls a ‘non-Humean’ theory of causation. The questions that we need to address here are: Does Dupré outline an alternative to physicalism? If so, what is it? And finally is it any more credible than the positions criticised in the first half of this work?

5.1.2 Promiscuous realism

Promiscuous realism emerges from an attack upon a common position concerning natural kinds. According to this view, associated with the writings of Kripke (1970) and Putnam (1975), natural kinds are individuated by essences. The original Kripke-Putnam account suggests that an essence is a microphysical structure shared by all objects of that kind: the essence of water is, for example, its microstructure, H₂O. However, some philosophers while acknowledging the significance of essences are more liberal in what they will count as essential. An essence may be some property or set of properties which is necessary and sufficient for the membership of that kind. On this account there will be a great many more essences than just microstructures.

Essences often figure explicitly or implicitly in some kind of physicalist thinking. Either in the broadly reductionist idea that each science identifies a more

or less general set of kinds which bear some strong connection (reduction, supervenience, whatever) to the kinds at the lower level. Or, alternatively, in the more eliminativist or non-reductive idea that the very essence of everything is determined by some fundamental science, i.e. physics. The first option is indicative of a more liberal attitude to kinds; the second is more in tune with the Kripke-Putnam line that essences are physical microstructures.

Dupré rejects both positions. The promiscuous realism Dupré advocates, eschews the idea that there is one neatly arranged way to divide up the kinds of the world. Inasmuch as this involves a rejection of the thought that there is any privileged set of kinds, physical kinds say, out of which we construct other, higher-order kinds, then this is a version of vertical pluralism. If Dupré considers his argument to show that at some level of description there are many ways to divide up the same set of things, then it is a version of horizontal pluralism.

Dupré's argument that there are many and various ways to class objects into kinds can be divided into two strands. First he illustrates that the terms of ordinary language cross-classify objects into different kinds and that, in turn these kinds do not match up with those identified by the sciences. To borrow one of his examples: for us it is very important in culinary terms to distinguish between garlic and onions but biology makes no such distinction. In fact, according to biological classification these gastronomic staples are part of the same class as certain lilies. So everyday classification cuts across scientific classification. Dupré's text is rich with many other examples.¹ Nevertheless, the naturalistic foibles of modern physicalists give them grounds to reject the significance of any such argument. Ordinary language, unlike science, they will say neither does nor

¹ See Dupré (1996) pp.26-34.

attempts to pick out natural kinds; therefore Dupré's observations about common linguistic practice have no bearing on how the world is or how we take it *really* to be.² We need then, if we hope to persuade the physicalist community at large, to turn to the second strand of the argument, that there is cross-classification of kinds within the sciences. Dupré's test case is the classification of species.

Species are not natural kinds in the Kripke-Putnam sense of being individuated by some microstructure – on that everyone agrees. As was emphasised in the discussion of teleosemantics, Darwinist thinking has highlighted the importance of variation within species and their historical development. Everyone agrees, therefore, that there will be no set of microphysical constituents shared by all organisms one would wish to classify as members of the same species. Nevertheless, there still remains the possibility that there are unique criteria, in the sense of necessary and sufficient conditions, for species membership. Dupré considers three such criteria of demarcation: morphological, biological and genealogical (or phylogenetic), each of which he argues is either practically or theoretically inadequate to provide a unified account of species.

Dupré first considers morphological criteria (p.44-5), the traditional basis for biological taxonomy. Species on this account are grouped together by likeness. As Dupré points out, this method suffers a major theoretical and philosophical drawback – there is no satisfactory account (nor could there probably ever be one) of objective similarity. Moreover, there seems to be little or no motivation for this type of grouping from biology itself. So the morphological account seems both theoretically suspect and scientifically unwarranted.

² Two of Dupré's detractors, Wilson (1996) and Wilkerson (1993) make this point. It is though quite a radical position, suggesting that we be eliminative with regard to most folk classifications.

Using concepts from evolutionary theory, accounts of species grounded in biology have been suggested and the apparent arbitrariness of resting classification on some ill-defined notion of similarity thus removed. What Dupré calls the biological account (p.45-6) defines a species in terms of the actual or possible reproductive links between individuals and reproductive isolation from other organisms.³ The thinking behind this idea is that reproductive isolation explains biological diversity. But here as well there are problems. One immediately apparent drawback for this way of taxonomising is that it can have no application to asexual species “everyone of which is isolated from every other.”(p.46) Moreover, Dupré argues, even within sexually reproductive species, isolation, particularly for plants, is not a condition that seems to be generally satisfied. One need think of the innumerable hybrids to see this. It seems unlikely, therefore, that reproductive isolation can be the sole criterion that explains divergent evolution and thus speciation.

The other biologically motivated alternative that Dupré considers, phylogenetic taxonomy, suffers limitations too (see pp. 47-49). Proponents of this particular view insist that a necessary condition for being a member of the same species is common descent from some set of ancestors. Subtly different approaches within the phylogeneticist school of thought will suggest additional criteria to provide necessary and sufficient conditions for species membership. The most radical view, cladism, maintains that species membership must be matched up exactly to the branches in the genealogical tree. This results in the radical claim (by taxonomic standards) that birds and reptiles should belong to the same class, since we believe both have common descent from earlier, primitive

³ Dupré quotes Ernst Mayr as the authority on this view of species.

reptiles. But whatever the particular approach, the general idea has some serious practical problems. It is very likely that, as Dupré says, genuine divisions by phylogenetic standards can be found well below the level of species. So the strictest of phylogenetic classifications may have to realign a great deal of commonly used classification. In itself this is not a major theoretical problem but, as Dupré points out, it is likely to leave a classification which is useless for many biological purposes. Moreover, there is no *a priori* reason why any monophyletic accounts should not come into conflict with groupings made on other equally well theoretically motivated grounds. Again to borrow an example of Dupré's, it may be that certain creatures that are genealogically divergent share the same niche. Therefore, from an ecological point of view it may be more useful to classify these organisms together as one species. Why, as he says, should considerations of monophyly override this?

Dupré makes a convincing case that if one looks at how the term species is used in the biological sciences today, there is no unique set of criteria on which that use is grounded; and, moreover, that any move towards a unique classification would be undesirable for biology. So we are led to the conclusion that classification of species, and classification in general, is not a matter of discovering a unique taxonomy "since there is none such; but [finding one] that serves some significant purpose better than the available alternatives." (p.52)

Why should the physicalist feel threatened by Dupré's observations?⁴ One possible line of thought might be that if different criteria can be used to pick out

⁴ One reason for thinking any species concept might threaten physicalism is that species appear vague; and that would create obvious problems when it came to trying to state physical bases for species. However, this does not appear to be any part of Dupré's argument. See Kenneth Waters (1998) for an account of species as vague.; and Brian McLaughlin (1997) for an account of how vagueness may be dealt with by a physicalist. I shall not discuss this particular problem however

different objects as belonging to different kinds, then there can be no objective answer to the question to which kind something belongs. Put like this, Dupré's argument will sound to some more like it is directed against *realism* about species rather than physicalism.⁵ Nevertheless, Dupré explicitly denies that he is anything other than a realist: "Certainly I can see no reason why commitment to many overlapping kinds of things should threaten the reality of any them." (p.262). One way to reconstruct Dupré's thought that might assuage such critics' fears would be to apply pluralism about taxonomy in a piecemeal fashion. It might be argued that certain criteria are more appropriate for grouping together some types of object under one kind than others; morphological criteria with asexual species, for example. In that case there always is a determinate answer to the question whether any given individual belongs to a particular species or not; it is just that the way that has been decided will vary from case to case.

Whatever the merits of such view it is clearly not Dupré's: "an organism might belong to one kind defined by a genealogical taxonomy and another by an ecologically driven taxonomy." (p.58) This suggests a different interpretation: each of the different classifications corresponds to a different science. That is to say, species is an ambiguous (as opposed to vague) term that has separate meanings in the separate biological sciences; certain types of classification may be more appropriate for ecologists and others more useful for, say paleobiologists. Framed in this particular manner, the claim seems far less radical than it may have appeared at first glance. If all Dupré is saying is that there are different classifications for the same object, all of which may be in some sense legitimate,

since it depends as much on your view of what kind of problem vagueness is as on any account of physicalism.

⁵ I think Wilson(1996) is tempted to see the matter like this. He suggests that the problems Dupré highlights only indicate that we have not quite hit on the right (or real) criteria for species.

that is surely uncontroversial. Consider, for example, a universe in which there exist three objects: a red square, a blue square and a red triangle. There are two systems of classification, corresponding to two sciences, the colour-science and the shape-science. The red square is – surprise, surprise – an object which belongs to two sets of classification, the red-types of object and the square-types of object. But it remains an open question whether the sort of properties characteristic of the red square are what they are in virtue of microphysical entities.⁶ Clearly something more than the idea that different classifications are possible is required if Dupré's argument is going to threaten physicalism and offer an alternative pluralistic ontology.

Certain remarks of Dupré's do indeed suggest he is offering something more than platitudes about classification:

The point rather is to make it clearer what should be the grounds for accepting a taxonomic scheme: not that it is the right one for there is none such; *but that it serves some significant purpose better than any available alternatives.* (p.51-52, my italics)

And more explicitly:

I have argued ... that the theories that science comes up with must depend on (that is will be relative to) the purposes for which they are intended. (p.261)

⁶ Dupré (1996) claims elsewhere that certain sciences, e.g. ecology, may simultaneously employ different concepts of species. Even if this is true, and I see no reason to deny it, there must be constraints on how species terms are used on pain of inconsistency. For example, I doubt Dupré

Dupré's thought might be that all classification *essentially* involves human assessments of purpose and significance. Therefore, no classification can be said to be simply supervenient on some base which ignores those interests and purposes; and unless those interest and purposes themselves could be explained away no kind of metaphysical reduction would be possible. This is an interesting (and I think not uncommon) thought.⁷ But Dupré fails to develop it further; he fails to explain why we must take classification *necessarily* to involve human interests and purposes; and indeed what ontological significance this has, if any. Many will think that the idea that classification cannot be divorced from human purposes is a straightforward admission of anti-realism. (Or if there are human interests involved, it can only be the desire to discover the *true* classification.) I do not discuss this topic here. Some of these issues will be dealt with in the next chapter. The question that concerns us is not whether Dupré is a realist or anti-realist about species, but whether promiscuous realism is a genuine alternative to physicalism.

Even if Dupré is claiming that all *biological* classification ineliminably involves human interests, I do not see any reason why there should be a conflict between the two doctrines. In fact, physicalists are in a strong position to turn Dupré's pluralism about classification into an argument for monism. If there are many ways of classifying individuals, the *same* individuals, we need some criterion for saying that the objects classified under one set of criteria are the same as those classified under another. What could this be? For the physicalist the

would wish to argue that an appropriate ecological analysis can use one notion of species one day and another the next as Quine has sometimes suggested one could about the physical world.

⁷ Arguably this is what Putnam was hinting at when he said that "the mind and world jointly make up the mind and world".

answer is simple. The same object is marked out by the physical properties from which all its other properties flow. Given an interest in specific properties for certain scientific purposes, whether it be size or colour or whatever, different classifications will emerge. Hence, the idea that classification of *non-physical* (or non-fundamental) properties essentially involves human interests can easily be accommodated by the physicalist. Unless the pluralist can extend Dupré's argument to show that it must apply to physical kinds as well, the interest relativity of classification should be of little concern. Physicalism and promiscuous realism about species are perfectly compatible.

Two important conclusions follow from these observations. First, Dupré has a problem about how to make sense of the notion of *same* object being *differently* classified. Second, and more importantly in the context of trying to develop a pluralist alternative, if promiscuous realism is compatible with physicalism, then we do not even have the beginnings of an argument for an alternative pluralist ontology.

Further investigation of Dupré's work makes it tempting to construe some of his remarks as revealing a commitment to a broadly physicalist ontology. For example, he advocates what he calls *compositional materialism* as a means of distancing himself from Cartesian dualism and other 'disreputable' metaphysical positions (pp.93-4). This weak kind of materialism commits one to the view that all concrete things are composed of microphysical entities. That is to say, he is committed to the idea that everything that is made of something, is made of physical things. Moreover, even when stating his pluralism, he is prone to comments which sound reductionist in tone:

The strongest possible notion of a real essence would be that of a property, or a group of properties, that determined ... all the other properties and behaviors of the objects possessing them... [S]uch a notion cannot work for type-essences, which are my present concern, for the simple reason that there are no kinds (*with the possible exception of microphysical kinds*) the members of which are identical to each other in respect of all their properties... (64-65, my italics).

The sort of physicalist who is relaxed about promiscuous realism for higher order kinds may only require essences, in the sense described above, for microphysical kinds. All other methods of classification are interest-relative, more or less so, depending roughly speaking, on how far they deviate from the basic microphysical kinds. Such a physicalist could readily endorse Dupré's comment that: "naturalness of kinds will turn out to be a matter of degree, some kinds will be a good deal more natural than others." (p.63) The more natural a kind the less anthropocentric it will be and the closer to the completely natural kinds of microphysics.

The following physicalist gloss on Dupré's views becomes tempting: If the world is fundamentally composed of physical objects, as Dupré suggests, then what such objects do must surely be a result of physics too. At base there is a set of physical objects and these physical objects have certain causal powers, completely specified by their essences. Just as the structural properties of the microphysical account for the structural properties of non-physical objects so too do the causal properties account for the large-scale causal properties of the non-physical. After all, they are just more properties.

It is the second strand of Dupré's pluralism which is supposed to block (or at least discourage) this kind of thinking. Dupré maintains that what things do is not simply a consequence of what things are. He advocates, what we might call, an *is/does distinction* in ontology.⁸ What a thing is made of, is just a matter of physics (or physical things); what a thing does, is not. It is this thought which is at the heart of Dupré's pluralism:

I ... suggest that less fundamental importance be attached to the delineation of kinds and attention rather be directed towards properties, dispositions and forces ... [W]hat makes a kind explanatorily useful is that instances share the same properties or dispositions and are susceptible to the same forces. (p.80)

The classification of kinds for Dupré comes *after* the discovery of properties, dispositions and forces; and this fact about discovery has ontological significance: properties, dispositions, etc. are fundamental. In other words, they are what some philosophers mean by the term 'real' – they exist independently of our organisational proclivities and ground the very possibility of such organisation. So here we have an answer, or at least the beginning of an answer, to the first problem I posed for Dupré. The identification of individuals does indeed seem to proceed on the basis of physical characteristics but he maintains that there is no way to go from knowledge of structure to knowledge of behaviour; and what objects do, their causal dispositions, grounds our classification into kinds; so those classifications cannot be reduced to or explained away by purely physical kinds.

⁸ His way of putting it is rather more Aristotelian: "why should we emphasize matter so strongly to the exclusion of form."

5.1.3 Causal pluralism

If promiscuous realism by itself is compatible with physicalism, as I have shown, then the key argument for Dupré's *metaphysical* pluralism is not that there are many ways to organise the things of the world but that the properties and dispositions of non-physical entities are not simply the result of their physical structure. What we need to try to understand here is how, if at all, this leads to an interesting and plausible alternative to physicalism.

Dupré's argument in favour of non-reducible causal powers again has two parts: one based on the failures of reduction in the life sciences and the second a metaphysical argument about the nature of causation. The first set of arguments will not concern us here. Most modern physicalists are happy to admit that there will not be a complete reduction of all disciplines to physics. I have, of course, argued that the ways physicalists have tried to do this are in general inadequate; that is the basis of the argument against physicalism in the first section. But no positive metaphysical alternative can be gleaned simply from these observations. If Dupré is to answer Melnyk's challenge, he needs to say something more constructive.

The basis of that more positive alternative is what Dupré calls a 'non-Humean' theory of causation. By 'non-Humean' Dupré means that causation or causal powers are to be taken as basic, rather than conceived of as constituted or grounded by regularities as, according to philosophical folklore, Hume claimed. Regularities, such as they are, are just one form of evidence for the existence and working of these causes. Dupré thinks this is a significant move because "[c]ertain views about the nature of causality suggest that only some kind of reductive

relation between higher and lower levels can achieve such consistency.”(p.99)
Here, he is clearly alluding to something like the overdetermination argument discussed in chapter 4.

Dupré seems to believe that by embracing this ‘non-Humean’ alternative we can avoid these implicitly reductionist commitments and understand the relations between the sciences in a new and pluralistic fashion. However, there is nothing about this ‘non-Humean’ position which is intrinsically inimical to physicalists. Indeed, I argued in the previous chapter that one cannot adequately formulate a core physicalist principle – the completeness of physics – employing only a regularity view⁹. So again it seems as though the elements Dupré is drawing upon are quite compatible with some kind of physicalist position. Nevertheless, even if the general elements are compatible with physicalism, it is possible to extrapolate something like an alternative ontology from Dupré’s remarks. He seems committed to something like the following: what is, is physical stuff and what does, does so in a way which is irreducible, even in principle, to what is or anything else for that matter (see in particular p.106 & 117-118). To put the same point less enigmatically, questions about structure can be answered using physics but questions about behaviour have their own autonomous explanations; and that reflects not lack of knowledge but the underlying metaphysical structure of the world.

I think that this represents in some way or another what Dupré is aiming for but is it any better than physicalism? Well, Dupré, if I have him right, will argue that it is superior because it is truer to the disunified state of current science. But there are conceptual problems that might make us suspicious that answer is a

⁹ Though, I think the non-Humean does create profound difficulties (or yet more profound difficulties) for versions of physicalism formulated using supervenience.

little glib. In particular, any advocate of physicalism is going to ask what the relation between structure and behaviour, between is and does, is meant to be on Dupré's account. Dupré embraces compositional materialism in an attempt to distance himself from Cartesian dualism but it could be argued that by opening up the gap between is and does, he creates another dualism which it seems equally hard to bridge. Without an explanation of the relation between is and does, physicalists might rightly feel that Dupré's metaphysics no better reflects the disunified state of science than their own. After all, they both agree (at least the non-reductivists among them do) that science as it stands now is, in fact, disunified. But it is not clear that either the physicalists or Dupré have provided a plausible metaphysical story to tell with which we can understand this disunity.

Let me reiterate: this is an extrapolation. Nowhere does Dupré explicitly commit himself to the metaphysical picture I have outlined. It is, however, the only *metaphysics* I can find in his writing that looks anything like a real alternative to physicalism; and Dupré is quite clear that he is putting forward an explicitly metaphysical account. His book is, after all, subtitled *Metaphysical Foundations of the Disunity of Science*. If the picture I have outlined is not what he means, then I do not know what is.

5.1.4 Concluding remarks on Dupré

What Dupré does brilliantly in his book is articulate the conceptual complexity of contemporary biology; and this I consider both highly important and informative. It reinforces the disunified picture of science that any who consider themselves remotely naturalistic in their philosophical leanings must accept. But I can't offer up any more than two cheers for Dupré. He fails to come up with a viable version

of metaphysical pluralism. The self-consciously different aspects of his world-view, promiscuous realism and non-Humean causation, do not appear to threaten physicalism. His commitment to compositional materialism on the other hand, either seems to lead to a certain kind of physicalism or what looks like an untenable distinction between is and does. These negative points are not surprising. Since Dupré concentrates on articulating disorder in the biological sciences and since he still seems, if only in a weak sense, to give physics a privileged role, he will always be vulnerable to the objection that there is either some mysterious relation between the physical and the non-physical or that if understood properly physics will be sufficient to explain away the apparent disorder in the other sciences. Of course, such objections are a long way from conclusive arguments in favour of any other alternative but there does seem to be a tension of some sort between metaphysical pluralism and compositional materialism.

The way to progress from this dialectical stand-off would be to see if some of Dupré's ideas could be applied to physics as well. In the following I shall discuss Nancy Cartwright's work which sets broadly similar ideas of Dupré's about causation in the context of the physics. We shall have cause later to see if some of Dupré's remarks about the interest relativity of classification and the consequent horizontal pluralism can be generalised when considering the works of Goodman and Putnam. My final position, I must confess, will in many ways be remarkably close to Dupré's. However, I hope to be able to provide a structure which can avoid some of the awkward questions Dupré's avowedly metaphysical pluralism inevitably raises.

5.2.1 Nancy Cartwright and the dappled world

Nancy Cartwright's attack upon physicalism or what she calls fundamentalism emerges from an analysis of the explanatory methods of physics; an analysis which foists upon us a different conception of causation and which forces us to reconsider the scope of our scientific laws. There are several strands of argument which promote a pluralistic metaphysics in her work: arguments based on analyses of causation and laws; on evidence; on the constructivist aspects of science; and on the interpretation of supposedly fundamental laws. The strategy will be the same as in the previous section: to see how far Cartwright's views really do force us to be non-physicalists and what arguments she offers to support an alternative ontology. I will begin with a general account of Cartwright's epistemology of science; from there we will be able to see how this is connected to her metaphysics and see what, if any, support it lends to her position.

5.2.2 Lying laws and honest capacities

The laws of physics lie – or so proclaims the title of Cartwright's (1983) first book. Why? Because in general the laws of physics hold, if they do at all, only *ceteris paribus*, that is other things being equal. To take a simple example: consider Coulomb's law for the force between two charged particles ($F = q_1q_2/4\pi\epsilon_0r^2$). In any real situation with two charged particles, separated by a distance r , the force will never equal that predicted by the law. All sorts of other factors will generally interfere – gravitational, magnetic and frictional forces, for example. Only if we have separate knowledge of and control over the interfering factors can we make the law come out true; and thus we can only have evidence that the law holds true for these particular circumstances. And, more importantly,

we already know that given less fortuitously arranged circumstances the law will be false. Hence in general the laws lie.¹⁰

To many this argument will seem strange. The general claim, so the thought might go, is not that the laws themselves simply describe relations between two kinds of directly measurable or perceptible properties. Rather, they describe typical kinds of effects and interactions which take place in the world and, obviously, one has to arrange the world in certain ways in order to make such effects transparent. Cartwright would no doubt agree. However, if one begins to think of laws like Coulomb's law as *describing* typical sorts of *tendencies* or *dispositions* that objects have, rather than as stating a general fact about the world, then one has already abandoned, according to Cartwright, one of the core ideas of empiricist philosophy of science – namely, that laws are regularities. What are really fundamental in this description of the world is not the laws but what Cartwright calls the capacities which we think they (sometimes) describe. Capacities or Aristotelian natures as she sometimes calls them (1999, ch. 4) are highly generic and open-ended dispositions or causal powers; for example, charge bodies have the capacity to attract oppositely charged bodies. Traditional forms of empiricism in the philosophy of science have taken talk of capacities or causal powers as the invocation of something mysterious and prescientific which should not feature in a mature and empirically tough-minded theory of the world. Cartwright, however, stands this idea on its head. Capacities are essential to the practice and understanding of the scientist. It is only those protected from the

¹⁰ The complete story is far more complicated, involving explanations of how theories via models are related to real world situations. The lying involved then becomes more dramatic than I have suggested above. It is not simply a case of the law truly applying in very limited real world situations, the law may not apply to the real world at all, only to the model used. The clever bit in science, according to Cartwright, is to match up real world situations to mathematically tractable models so we may obtain useful results and explanations.

complexities of experimental science who would think otherwise. What justifies the enormous amount of money spent on research, if anything does, is that scientific theories and knowledge can be exported beyond the walls of the laboratory to have new application. If we consider laws to be basic we can, Cartwright argues, have no justification for thinking that the results of physics hold anywhere else in the world – these laws hold, if they do at all, only *ceteris paribus*. Only if we consider laws, like Coulomb’s law, to describe tendencies to produce certain effects (even if they are frustrated in various ways) can we expect our knowledge gained in the laboratory to tell us anything about the world outside.

My immediate interest is not so much in the quality of this argument but its bearing on physicalism. How does Cartwright turn an argument in favour of capacities into an argument in favour of pluralism? I think her writing is a little confusing on this matter. In the next few pages I will try and highlight some blind alleys that Cartwright’s philosophy might lead, before sketching some more promising alternatives.

5.2.3 Making laws

Laws are made, says Cartwright (1997, 1999, ch.3). We make laws like Coulomb’s law by arranging the features of our experiments so they will be true; we *shield*, we *interfere*, etc. – we *make* what she calls *nomological machines*. This making presupposes capacities. They are, so to speak, the ingredients in our nomological recipes.¹¹ Let us agree that it is so. Cartwright needs to tell us more

¹¹ Cartwright (1999, p.77) claims that although capacities are basic, she means this to be taken in neither an epistemological nor metaphysical sense. The latter exclusion is odd, I think. The reason she offers is that law or regularity claims are as true as claims about capacities. Of course, this can be acknowledged while still giving metaphysical precedence to capacities; any reductionist story

about these capacities. As a self-proclaimed empiricist what must be said about capacities has to be related to experiment and measurement if they are to be a respectable addition to our ontology. According to Cartwright, capacities are empirically kosher because they are, in certain, circumstances, directly measurable.¹² But that is not especially the part of Cartwright's story that interests me here; it is the conclusions about the world we can draw after we have done our experiments and made our measurements that I want to discuss.

Cartwright maintains that when we are not able to control things in our laboratory, when we can't build nomological machines¹³, we can't apply our scientific knowledge. She provides a simple example of an experiment that defeats theoretical treatment. Imagine we were to drop a \$1000 bill from a window on a blustery day. Away from the neatly arranged experiments of the laboratory we have no way to begin to model this situation with our current physics (or certainly at least not our elementary mechanics). Since even this mundane example is not covered by some nomological machine and it is only where we have such machines that we have laws, Cartwright argues, laws are just not universal.

These kinds of examples are not apt to impress physicalists. Everyone knows, they will say, that we have no strictly accurate model for this situation. The reason why we do not, though, is equally obvious – mathematical and practical complexity. We can't model the situation not because there is something wrong with our physics but because either we do not have enough information of

will say something similar. Since it is hard, if not impossible to see what else Cartwright might mean here, I take it that we can safely say that more basic is, in fact, meant in a metaphysical sense and we can say that without impugning the reality of regularities.

¹² This part of Cartwright's theory of capacities is developed in detail in Cartwright (1989)

the initial conditions or the mathematical know-how to say anything useful. For Cartwright though (and for me also) this response is far from obvious. Why should we believe that our inability to model this situation using classical mechanics is *only* due to problems of complexity? Where is the *evidence* or the *argument* for this claim? One, of course, could write down various systems of equations that one might think captured certain aspects of the interaction, and show that they were mathematically intractable. For example, one might think of the wind in terms of lots of little particles bombarding the bill and try to write down some highly complex and mathematically intractable force function. But so what? That only shows it is possible to devise mathematically intractable equations; it does nothing to show that this is the right force function. Only if you already believe in fundamentalism would you find any such arguments compelling. Without a treatment of this supposed force function that actually can predict the motions of the bill,¹⁴ we have no reason to think that the wind can be described by any force function.

The negative lesson, very familiar by now I hope, is that physicalism is not well supported by scientific practice or evidence. But what positive conclusions follow? Cartwright suggests we take our laws were we find them (or perhaps more accurately, where our nomological machines build them). The world is made up of lots of laws which apply in lots of different circumstances; sometimes overlapping, sometimes running out altogether. Perhaps we can't get an explanation for our \$1000 bill in terms of the laws of Newtonian mechanics but we might do better with fluid dynamics; and perhaps in some cases we will have

¹³ I should point out that Cartwright does acknowledge that certain nomological machines are naturally occurring – the planetary motions are an obvious example. However, this is the exception rather than the rule.

nothing worthwhile at all to say on the matter. We have what Cartwright calls a patchwork of laws, rather than the hierarchy dreamt of by physicalists.

Cartwright calls her position ‘cross-ways’ emergentism or pluralism and the name is apt for it cuts across Price’s distinction between horizontal and vertical pluralism. Price’s distinction seems to presuppose some kind of hierarchy which we can move up or down. However, Cartwright’s ontology allows for many laws, at the same level of organisation, each true in a certain domain of application – so it cannot really be classified as vertical pluralism. On the other hand, it does not invoke the idea that the same things may have different descriptions, rather just that various sorts of description are limited in their applicability. So we have an ontology which is thoroughly pluralist and yet escapes being classified as either horizontal or vertical pluralism. Indeed the whole idea of there being any particular way to organise the structure of reality (apart from trivial notions like size) is lost. We work with what works and that is all.

Be that as it may, Cartwright’s metaphysical conclusions seem a little queer in light of her claims about laws and capacities. Remember the world is a world basically of capacities, not laws; laws are made using capacities. Moreover the main reason for thinking that ours is a world of capacities is that capacity claims can be exported into new contexts, whereas laws only hold *ceteris paribus*. The question arises then does Cartwright understand the patchwork of laws as a patchwork of regularities or as a patchwork of capacities? Consistency of usage would seem to demand that she means regularities.¹⁵ If that is so, then it is

¹⁴ Of course such a description should not bring any elements external to the supposedly fundamental theory. *Cf.* my discussion of quantum chemistry in chapter one.

¹⁵ Cartwright’s pluralistic metaphors also invite this interpretation. Talk of a “dappled world” or a “patchwork” is suggestive of an impressionistic and superficial level of description.

difficult to see what metaphysical weight her position would carry. Fundamentally the world could be made up of a few basic capacities interacting with one another giving rise, at the phenomenological level to a patchwork of regularities. In other words, a fundamentalist could accept Cartwright's arguments concerning capacities but insist that there are very few, fundamental capacities which account for the behaviour of everything. On the other hand, if Cartwright's observations about the limited power of scientific explanations apply to capacities, then the superiority of capacity claims over regularity claims looks threatened. If capacity claims hold only in limited circumstances, or one might be tempted to say, *ceteris paribus*, then they suffer the same faults as regularities. This particular dilemma is forcibly put by Margaret Morrison (1996) and reiterated by others (see for example Rueger and Sharp (1998)).

One way to relieve the tension might be to accuse Cartwright of sloppy writing. When she has used the term law, sometimes she has meant good old empiricist regularities and sometimes she means what we might call 'real laws' – generalised capacity claims or something similar. Real laws have *greater* generalisability than do regularities but they are not completely universalisable. Rather, they will have a certain domain of application. This interpretation is suggested in Cartwright's (1996) response to Margaret Morrison and is supported by other remarks Cartwright makes (pp.53-54) 1999)¹⁶ which interpret the laws of physics as describing capacities. Assuming it is right to understand her as referring to capacities, we do not yet have a metaphysical argument that capacity claims only hold within certain domains in the way we do for regularity claims or for the application of nomological machines. A physicalist could agree with

¹⁶ Possibly her remarks (1999) on p.29 support this: "It seems to me wholism is far more likely to give rise to *ceteris paribus* laws, whereas natures are more congenial to pluralism."

everything Cartwright has said about nomological machines and capacities but still insist that there were only a few *fundamental* capacities that explained the patchwork of laws.

5.2.4 An interlude regarding causation

Cartwright, like Dupré, is offering a non-Humean analysis of causation as a first step to undermining physicalism. I must repeat that I am sceptical that any such analysis can aid matters at all. As I have argued above, it is not at all clear that one cannot formulate a coherent version of physicalism in terms of causal powers or something similar. However, I feel I should pause to say something more about causation before moving on.

I agree with Cartwright and Dupré that a Humean account of causation is wrong. My reasons are boring and familiar: the problem of accidentally true generalisations, that there is no privileged set of properties (impressions, sense data, occurrent properties) from which to fashion regularities. But I do not have an alternative analysis to offer. In fact, I do not see much need when talking about causation to analyse away our common understanding. I cannot see what pressure there is to offer such an analysis unless one believes there is some privileged set of properties (sense data, occurrent properties... all the usual suspects) from which everything else must be constructed. And I do not see why anyone would believe that there were such special properties, unless they were beholden to an empiricist epistemology; and I do not know of any remotely plausible argument which suggests why we should be so beholden.¹⁷

¹⁷ There are ways to link these ideas conceptually with physicalism. That is to say, one might define occurrent properties as the basic physical properties of the universe. Such a stipulation, however, can hardly advance the case for physicalism; nor its denial promote an opposing ontology.

This is not to say nothing can be said about causation. For example, causation must, I think, be admitted to have a modal and a lawlike dimension. That is to say, causes really do bring about their effects in some way. It is part of our conception of the fact that one event caused another that the same type of event would cause such an event again in similar circumstances. This is of course to offer up platitudes not analyses – I use the terms ‘lawlike’ and ‘modal’ with unashamed informality. Causation is such a broad and highly abstract concept that it seems to me that we cannot hope for much more here. If we must have more, then I would follow Donald Davidson (1995) in urging that the way we pick out objects, their causal dispositions, and how these dispositions relate to each other and work together is a largely holistic project. So our very conceptions of causation, event and object are all intimately bound up with one another. Davidson is right when he says that: “events are changes that require explanation. This is not an empirical fact, we decide what counts as a change on the basis of what we want to explain and what we think available as an explanation.”(p.273) Davidson’s point tallies well with Cartwright’s and Dupré’s insistence that one cannot have knowledge of what something will do unless one has knowledge of the context. Being able to identify a cause in context requires that one can understand what is and what is not a change and what requires explanation. Indeed, this is why I do not think you can possibly draw an interesting distinction between is and does in the way that Dupré attempts. Such divisions, when we make them, reflect our explanatory interests as much as any others. One can see this very clearly at the more formal end of physics. Is and does are reflected in divisions into “spaces”, not necessarily physical ones, and interactions within those spaces. These divisions provide the background against which we discuss

and label changes or events, broadly construed. Where I would demur from Davidson's analysis, is in his insistence that there is an a priori reason to believe that there is one framework in which it is possible to formulate the laws in terms of some complete theory or other. The way we approach the formulation of theories is piecemeal and this is reflected in the conceptual mess of contemporary science. But there is no reason to suppose that there must be some way to phrase things or parse things which will provide in any interesting sense a complete theory. The difficulties of relating quantum mechanics and classical mechanics illustrate the problems well enough.¹⁸

These are little more than opinions, of course. I do not expect my observations regarding causation to persuade those who take a different view. The point is to reiterate how little an account of causation will contribute to whether or not one is a physicalist. It is the relation between types of cause that matters. Davidson and I are largely in agreement about causation but not about these relations, as will become apparent later.

5.2.5 Wholism and pluralism

Capacities understood as an alternative to Humean causation do not lead to pluralism. Nevertheless, there are other features of Cartwright's capacities that are more promising. Before moving on to discuss these issues, it is worth looking more explicitly at Cartwright's writings on metaphysics. What she says here leads to another dialectical dead-end that it is worth illuminating so we may pass safely by on our way to constructing a viable pluralism.

¹⁸ Formally, one might consider the problem of quantum measurement as how to relate changes that occur in a Hilbert space, with changes (or measurements) that do not. See Albert (1996) for a

In an interesting article, Cartwright (1991) offers a third alternative to physicalism and pluralism – what she calls wholism [sic]¹⁹. To motivate such an ontology we begin, again, with the observation that our laws are formulated in highly specific circumstances in the laboratory. We shield and protect the objects we are interested in from external influences as far as possible. The wholist believes that the way we discover such “laws” may systematically mislead us about how the world is. The interactions studied by us are just small parts in an intimately connected whole. Wholists believe the behaviour of entities is dependent on the total structure, rather than the individual behaviours of the parts. It would therefore be unwise to draw wide-ranging conclusions from our narrow experiments.

Wholism is more than just fancy for Cartwright, she gives an example from physics that might lend support to such intuitions. The ability of fibre optic cables to carry information with little loss is the product of two oppositely acting factors. All fibre optic cables suffer from pulse broadening (that is the light signal becomes more diffuse as it travels along the fibre) due to dispersion effects associated with all optical phenomena. However, this effect is counterbalanced by pulse narrowing effects. The intensity of the light causes a small shift in the refractive index of the fibre, producing a non-linearity. This creates what is known as a ‘chirp’ which is essentially the pulse-narrowing phenomena which cancels out the dispersion effects. Hence one gets a stable, “soliton” wave with virtually no information loss. And that’s just as well for all of us since that allows us to move large amounts of information about very efficiently. Cartwright suggests two interpretations of situations like this. One is her standard capacity account

discussion along these lines.

¹⁹ I will use her spelling throughout (and drop the sics too).

which relies on stable tendencies. The other is the wholist picture which denies there is any underlying stability which would ground inductions from one case to another. The soliton phenomenon is a fortuitous result of the total arrangement of the universe. On the wholist picture, there is no way to analyse the phenomenon in terms of the stable interactions of parts.

Ultimately, though, Cartwright rejects wholism because she claims “metaphysically the fundamentalist is borne out.”(1999, p.31) Since the success of science is to be explained in terms of some fundamental arrangement of the universe, some general theory of everything, Cartwright thinks that is just another version of the fundamentalist creed. But unless one holds *a priori* that there is something wrong with fundamentalism, this does not seem a compelling reason to reject wholism.

The mistrust of this kind of general, high-level account that Cartwright expresses here is explained by her empiricism. What Cartwright believes is really wrong with wholism is that, like the other forms of metaphysical fundamentalism, we have no evidence for its truth. As she says: “I am prepared to believe in more general theories when we have direct empirical support for them.”(p.31) No evidence, no metaphysical commitment. All that is left is, as was hinted at earlier, to take science at face value. But that might mean one of two things – either we abjure metaphysics all together or we build our ontology from our best confirmed science. Cartwright takes the latter option:

Metaphysical nomological pluralism is the doctrine that nature is governed in different domains by different systems of laws not necessarily related to each other in any systematic or uniform way: by

a patchwork of laws. Nomological pluralism opposes any kind of fundamentalism. (p.31, my italics)

In light of the discussion above, we should, I maintain, interpret laws as here meaning some generalised capacity claims. Cartwright's position, then is that this kind of pluralism is the metaphysics best supported by the practice of the sciences.

In my view she takes the wrong option. The argument used by Cartwright against all other metaphysical positions can also be used against her. What is the evidence, the *direct* evidence, that the universe is made up of a patchwork of laws? Perhaps it is true that we have good evidence for each of the laws that we take to hold true but that is quite different from having evidence of the *general thesis* that the world is made up of a patchwork of laws. In acknowledging that both physicalism (1993) and wholism (1991) are possible ways to construe the world, Cartwright has only provided one other possibility, one other metaphysics. What *direct* evidence is there that this *metaphysical* position is true? No more than for the other two positions she rejects.

It is difficult to understand why anyone as empirically minded as Cartwright should be bothered by the metaphysics of the situation at all. If she does not see "why we need to explain [scientific] success" (p.31), then what role has metaphysics to play? The patchwork view of the world just seems to be another explanation of that success, except it is one which in addition explains why we should never expect to obtain a theoretically neat conception of the world. If one insists that one should not commit oneself to anything other than that which can be supported by the best scientific evidence, then one should simply abstain from any kind of global theorising about the nature of the world.

Cartwright it seems to me has a telling (and in this work familiar) argument against physicalism – it is not backed up by the practice of science. But she has little to offer in these papers to answer Melnyk’s challenge. The arguments from causation and evidence fail to show how metaphysical pluralism is any more reasonable than physicalism. In fact, if Cartwright were to apply her evidential standards consistently, then her own favoured metaphysical position should be ruled out.

5.2.6 Capacities, laws and abstract concepts

I think there are passages in Cartwright’s writing that might allow us to circumvent some of the problems I have raised. In particular, a charming article (1999, ch.2) of Cartwright’s suggests a way of thinking of capacity or law claims that moves away (or at least can be interpreted as moving away) from the kind of metaphysical commitment implicit in the other articles I have discussed.

Cartwright argues that there is an analogy between the relation of high level laws, like $F=ma$, to concrete reality and the relation of morals to fables. Specifically, she thinks that laws involve *abstract* properties which can only begin to be judged true or false in the concrete context of application. Her ideas draw on the work of Gotthold Lessing’s account of the relation between simple morals and the fables that are often invented to illustrate their force. According to Lessing’s account, or at least Cartwright’s interpretation of it, “the relationship between the moral and the fable is that of the general to the more specific.”

The account of abstraction that I borrow from Lessing to describe how contemporary physics theories work provides us with two necessary

conditions. First, a concept that is abstract relative to other more concrete set of descriptions never applies unless one of the more concrete descriptions also applies. These are the descriptions that are used to 'fit out' the abstract description on any given occasion. Second satisfying the associated concrete description that applies on a particular occasion is what satisfying the abstract description consists in on that occasion. (p.38)

Now, this can sound like an edifyingly romantic reformulation of Cartwright's theory of nomological machines – laws hold only where we can carefully control the circumstances and weed out possible interference. But one can interpret this paper slightly differently, I think. The alternative interpretation, which I hope to develop and enrich in my discussion of Putnam and Davidson, would be to suggest that once we think of laws as abstract, it does not even make sense to think of them holding everywhere. That is to say, without detailed knowledge of the how to 'fit out' the law, we cannot even begin to say what it means for the world to be governed by a set (or the set) of fundamental capacities that we take these laws to describe. Wholism and fundamentalism then are not viable metaphysical alternatives but make senseless claims based on a misunderstanding of high-level laws. This shifts the argument against physicalism away from evidence and towards claims about the best way to understand our physical theories. In other words, away from offering up an explicitly alternative

metaphysics, towards an evaluation of the sorts of philosophical claims that are available.²⁰

I do not know if this is the gloss that Cartwright intended to put on her views but it seems to me to be an interesting and significant way to avoid physicalism. By considering some explicitly pragmatist theses, Cartwright's specific claim about high-level laws can be put in a context in which a viable version of pluralism can be defended. However, since the arguments are quite involved and Cartwright's writings offer only a hint as to how to develop such an alternative, this discussion will have to be postponed until later chapters.

5.2.7 Making entities

Another possible interpretation of Cartwright's pluralism is suggested in the closing paragraphs of her essay on the patchwork view of laws. Here she makes some interesting remarks regarding theoretical entities. These comments have a bearing on Dupré's is/does distinction and should allow us to segue neatly into a discussion of Nelson Goodman.

Ian Hacking is famous for arguing that ontological commitment to theoretical entities should be based on our ability to manipulate such entities to create new experimental phenomena. The argument is neatly summarised in the phrase: "if you can spray them, they exist". Cartwright wants to change this to "*when* you spray them, they exist". (1999, p. 34) The idea here being I think that we can only have sound knowledge of various phenomena in limited circumstances. Or to put in the language of capacities: we have knowledge of the capacities thought typical of various entities in limited circumstances, so we can

²⁰ Note the same ambiguities in interpretation arise here as elsewhere in Cartwright's work because she is far from clear about what the relations between theoretical laws, regularities and

only be committed to the existence of such entities in equally limited circumstances. Now this reiterates the argument for a patchwork of laws at the level of entities. Clearly and unambiguously, it also represents a denial of the kind of is/does distinction that I have attributed to Dupré. The upshot is a position which suggests that what there is, the world, is partly a product of our experimental enterprises. Cartwright (ibid.) acknowledges that her position is very close to that of certain social constructionists here, who consider all knowledge claims to reflect our ways of constructing rather than the world itself. Here, her overall position sounds less like a traditionally realist metaphysics and more like a rejection of metaphysics outright. In other words, more like the sort of position to which, I think, her arguments lead. Again Cartwright offers here only a hint at this anti-metaphysical alternative. One could develop this idea further by considering the social constructionists Cartwright mentions. Instead I wish to turn to the writings of an analytic philosopher, Nelson Goodman. Goodman develops in a rigorous way the idea that the world is constructed by us and that this inevitably leads to pluralism.

Chapter 6

From Pragmatism to Pluralism

The next two sections deal with two neo-pragmatists: Nelson Goodman and Hilary Putnam. Both philosophers could rightly be called anti-physicalist but this would hardly do justice to their views. Goodman defends a position which he calls ‘irrealism’; a philosophy which opposes just about all of the conventional ontologies in the canon. Putnam’s own position is ever changing and elusive both in name and content but one fixed point in his philosophy has been an attack on what he calls metaphysical realism. A position which he believes in its modern form is intimately connected to physicalism but also includes various forms of idealism and dualism. Because the targets of these two philosophers are much bigger than my own, it may appear that I am wandering from the path of constructing a pluralist position. But this is not so. In what follows I shall be concentrating on the special sorts of argument that Goodman and Putnam offer for *horizontal pluralism*. Both philosophers, but particularly Putnam, provide insights into the sort of position that one might credibly construct that would deserve the name pluralism; and it is on these insights I intend to build in the final chapters.

6.1.1 Making sense of Goodman’s worlds

Goodman’s writings are layered with metaphor, quips, puns, deliberate paradox and other devices, as he says, to keep the reader awake. In this he surely succeeds. But for those of us who pursue philosophy in a more prosaic fashion he leaves many questions and few answers. In what follows, I shall attempt to reconstruct sympathetically, carefully and, of course pedantically, exactly what Goodman’s views are.

6.1.2 Conflicting truths

Consider the following two sentences:

(A1) The Earth is at rest.

(A2) The Earth is moving.¹

(A1) and (A2) are contradictory, on that we can all agree. Goodman claims that both sentences are also true. But two contradictory sentences cannot be true together. Their joint assertion being a contradiction, anything would follow and that would obviously be a disaster. How is it then that two true but contradictory sentences can be asserted? The answer is simple, says Goodman, both statements are true but of different worlds; no contradiction arises because we no longer consider both assertions to be referring to the same world. This is a bold sounding way out of Goodman's little puzzle but what on Earth can it mean – surely there is only one world, a unique *universe*. Nothing could be more obvious than that. It is in determining what Goodman does in fact mean by a world, and how there can be more than one, which will mark out the path to our understanding of his pluralism.

6.1.3 How on earth can you make worlds?

Let us begin with the obvious. We make what Goodman calls versions. Versions are symbol systems² of some kind which are intended to communicate certain thoughts, beliefs, emotions and other such characteristics. Clearly if anything makes a symbol system then we do. So far, so obvious. And it of course then follows that if any versions that are made are true versions, then we make true

¹ See McCormick (ed.) (1996) pp. 80-82 for Goodman's discussion of this example.

versions. Goodman, though, then goes on to claim that true versions make worlds. What can this mean? What a lot of commentators seem to think Goodman means is that we make versions true. There is a sense in which I think this interpretation is correct – but more needs to be said. Essentially, Goodman offers three arguments (or better three stages of argument) to establish that we make worlds. One argument proceeds directly from considerations of what we do and do not make; the second from criticisms of the certain views about truth and the third from further considerations of two apparently true but jointly contradictory statements.

The best place to begin our unravelling of Goodman's first argument is an exchange between Goodman and Israel Scheffler. Scheffler, a perceptive and persistent critic of Goodman's views, challenges Goodman to illuminate what he means by saying that we made the world by asking did we make the stars. (McCormick (1996), pp.138-9) Goodman, boldly replies that we did indeed make the stars and invites the reader to consider the following dialogue to illustrate his point.

The Great Dipper [sic] was made by an adopted world version.

No, it was made by nature.

Did nature make it the Great Dipper?

Well no; it was made the Great Dipper by being picked out and so called by a version.

What is it that was made by Nature and was there to be picked out and named?

² Strictly speaking symbol systems are only one sub-class of versions; some versions like works of art are non-symbolic.

A particular constellation.

Was it made a particular constellation by Nature?

Well, no; it was made a constellation by a version that distinguished certain configurations of stars from other under the general term 'constellations'.

But did Nature make the stars?

Certainly

Did it make them stars?

Again no; they were made stars by a version that distinguished certain conglomerations of particles, or objects in the sky, from others under the general term 'star'.

Did nature make the –

This could go on and on; but your arguments seem at most to show that without versions stars do not exist *qua* stars, not that do not exist at all.

But do stars-not-*qua*-stars, stars-not-*qua*-moving and not-*qua*-fixed, move or not? Without a version, they are neither moving nor fixed.

And whatever neither moves nor is fixed, is neither *qua* so-and-so nor *qua* not so-and-so, comes to nothing. (McCormick (1996), pp.166-67)

The gradual transition in this dialogue from a concession about making a constellation (namely the Big Dipper) to making constellations to making stars is worth pursuing in greater detail. In what sense would we all agree that we make the Big Dipper? Certainly we all agree that we made the name "the Big Dipper" up and then applied that name by some linguistic function (pick your favourite

theory of reference) or other to those stars. So we made the Big Dipper, “the Big Dipper” by so calling it exactly that. But then, the question arises for Goodman, what part of the Big Dipper did we not make? And the answer is obviously we did not make the particular stars that constitute the particular constellation which we call by the name “the Big Dipper”. And of course, Goodman’s riposte is that in exactly the same way that we made the Big Dipper the Big Dipper, we made constellations, constellations and stars, stars. But this is too quick. The Big Dipper is after all a proper name, not a concept and we make things have proper names by so calling them that proper name.³ However, we do not make things fall under concept terms by so introducing concept terms. It is true that words like star, constellation, etc. have parameters determined by some linguistic convention or other. It is indeed also true that we in a certain sense make the concepts star and constellation (or perhaps we can put this more clearly by saying we invent the words “star” and “constellation” for the concepts they pick out). But it is equally obvious that we do not make any thing fall under those concept terms simply in virtue of introducing those terms into our language. It is (in some sense yet to be properly articulated) up to the world whether anything falls under these concept terms. Goodman’s argument as it stands seems to conflate our use of proper names and concepts to make his point.⁴

The foregoing analysis I believe is at least on some superficial level correct. It explains why Goodman’s claims sound so counterintuitive. But if the argument were so obviously defective then one might be left wondering how an

³ Of course, if your favourite theory of proper names is that they are in fact complicated quantifier expression, then not even Goodman’s claim that we make the Big Dipper, the Big Dipper can be supported.

⁴ Putnam gives precisely this analysis of Goodman’s argument in “Irrealism and Deconstruction” in both Putnam (1992) and McCormick (1996). A similar point is made by Michael Devitt (1997, p. 245).

astute philosopher like Goodman could ever have proposed such nonsense. The right answer to that question concerns (as always) truth – or at least, two conflicting notions of truth. Goodman’s argument implicitly assumes that there is something wrong with a certain conception of truth (and equally on one level the response that Goodman’s argument is nonsense presupposes this very conception); and that as one tries to illuminate that conception one will come unstuck. It is our unguarded use of phrases like “the world makes certain statements true” (or “the world decides whether certain concept terms are non-null”) that Goodman is concerned to show are problematic. His real worry is that our simple argument against making worlds relies on an undiagnosed use of a correspondence theory of truth. That is to say, if we press the case against the correspondence theorist and ask him to explicate how it is this world apart from any version makes statements true, he will be at a loss as to explain how and indeed what this world is. So, as I interpret Goodman, if the way out of his first argument is to be viable, then we need an analysis or some explanation of this version-independent world and in what sense it is already there. Goodman is explicit that this cannot be done and to prove his case makes some familiar observations about how much conceptual machinery we need to pick out what we might think “is already there”.

I sit in a cluttered waiting room, unaware of my stereo system. Gradually I make out two speakers built into the bookcase, a receiver and turntable in a corner. I find a system that was already there. But see what this finding involves: distinguishing the several components from the surrounding, categorizing by function and uniting them into a

single whole. A good deal of making, with complex conceptual equipment, has gone into finding what was already there. (McCormick (1996), p.155)

This can be read as making essentially the same point as the star-making argument; one that moves from saying we make concepts, or concept terms, to saying that we make worlds. It can though be interpreted as suggesting that the idea of finding what is already there, what a correspondence might mean by the world independent of all versions, is a non-starter for any kind of explanation of how we use and deploy concepts. Saying anything about this world-as-it-is-in-itself will necessarily invoke a version. To say anything at all one must use words and one cannot use words meaningfully without using a particular version. Now together with the first argument we can see how this observation makes the case for star-making. We suppose when we say that it is the world that makes certain objects fall under concepts, that this is unproblematic. But Goodman's second argument shows us that when we try to say how the world is, we will have to fall back on some version or other. So we will not have any sense of the world as-it-is-in-itself but rather some version of the world; that is some description of the world. So our claim that the world makes it the case that certain objects fall under certain concept terms true turns out to be empty, since all we can mean by world here is just some version or other. Given the impossibility of illuminating the concept of the world-independent-of-versions, Goodman suggests that the only way to define truth is in terms of some more general notion of rightness; and that notion will have to use version-dependent factors to provide a complete

explication. Hence, epistemic considerations like coherence, warrant, etc. must be used in the definition of truth. (See McCormick (1996), p.157-8)

One might summarise this particular line of thought by saying that either what is meant by the-world-in-itself is just some version or other, or else it is something about which nothing can be said. That is to say, the explanation of truth as correspondence to the world-in-itself either collapses back into a version or is explanatorily redundant. Such anti-correspondence arguments are common enough. However, as Donald Davidson (1990, pp.302-303) has recently pointed out, Goodman's argument and others like it are fatally flawed and question-begging. Why should any correspondence theorist or any other believer in the world-independent-of-descriptions have to be able to give an account of the world independent of a version? After all, his entire position rests on the idea that there is an extra linguistic reality which is quite independent of any versions or beliefs. To tell him that he is unable to provide any positive account of how this world connects with our beliefs and versions is not to present him with a devastating objection but simply to outline one of the basic claims of the correspondence theorist. Namely, that the way the world is, is independent of our beliefs.

Nevertheless, one may still be disturbed by Goodman's point. If the choice is between Goodmanian irrealism in which the world is made and the correspondence theory in which the world seems as though it may forever be beyond us, then some might feel it is better to have a world that has some definite connections to our beliefs or versions, however sullied by human making. Moreover, despite the popularity of talk of correspondence, no one has been able to come up with a remotely plausible theory that can cash out the metaphor. In particular, there seems little prospect of making sense of the idea of a fact as

something which sentences correspond to: problems with negative facts and Davidsonian slingshot arguments⁵ have demonstrated as much.

Fortunately, though, we are not forced to make a choice between irrealism and the correspondence theory of truth. A perfectly ordinary sense of the world or the objective can still be recovered that does not invoke any mysterious notion of correspondence. The alternative is very simple and merely requires one to pay attention to our use of language. The way we talk, our linguistic competence, presupposes that there is something to be talked about other than the version. Speakers are aware there is a difference between simply the word “armchairs” and armchairs; between, as philosophers say in the jargon, *mention* and *use*.⁶ And any and everyone who can use a language competently is aware of this difference. (Just as any and everyone who has actually been punched in the face is aware of the difference between being punched in the face and just talking about (or representing) being punched in the face.) One is quite entitled, therefore, to say that some version of the world is *about* the world and if one is asked to describe the world one may do so unproblematically *using* some version or other. Such using does not require any theory of correspondence to make sense of what is going on, only an ability to speak the language. The point is prosaic but worth repeating: being able to keep track of conversation requires us to know the difference between use and mention, between words and the world; and being able to note that distinction does not involve implicit commitment to a correspondence theory of truth or anything else which might be considered philosophically suspect.⁷

⁵ See Davidson (1984), “True to the Facts”.

⁶ This criticism can be found in Devitt (1997), ch. 13, Putnam (1992), p.122.

⁷ What looks like a deeper discussion of these issues can be found in an exchange between John McDowell (2000) and Richard Rorty (2000). McDowell insists that the idea of objectivity,

6.1.4 Conflicting versions again

Thus far we have seen that the first two stages in Goodman's arguments that we make worlds and worlds are just true versions are, by themselves, wanting. But without the third part of Goodman's argument – the idea that there can be jointly inconsistent but true sentences – the case for worldmaking is not yet complete. This is the critical stage in Goodman's argument. If there can be two true but mutually inconsistent statements, then the idea that there is a world which both these statements are about seems to be hopeless. Once one makes Goodman's move from one world to many, the conclusion that there is none at all follows quickly.

“[W]orld” is all inclusive, covers all there is. A world is a totality; there can be no multiplicity of totalities...By assigning conflicting versions to different worlds, we preclude composition of these totalities into one. Whatever we may mean by saying that the motion of the Earth, or of different earths, differs in different worlds, we rule

intimately related to the idea of the world, should not be given up. He claims that our very understanding of truth as involved in disquotation involves the norm of being “answerable to the world”. And that what McDowell calls claim-making requires that, as he puts it, “we direct our meaning” (p. 119) at the world. Again, this is in essence the same point as made above. It is part of putting a sentence forward as true, making a *claim*, that you *use* it to say something about the world. It *must* be thought of as more than a conversational manoeuvre in a game with you peers. Rorty responds by denying that any significance can be given to the idea of “answering to the world” as opposed to trying to achieve solidarity with ones peers. This is clearly the analogue of Goodman's implicit challenge to say what this world is without invoking versions. So the argument here moves in the same small circles. The important point to repeat, which I think Goodman and Rorty ignore, is that one is not forced to choose between solidarity and objectivity or true versions and the world. Rather the pursuit of solidarity, the serious pursuit of solidarity in which we want to get things right, not just stop arguing so we can go to bed early, also involves the idea that we are answerable to the world.

Rorty in his response to Bjorn Ramberg (2000) seems to accept something like this when he explains Davidsonian triangulation as follows: “It [is] a mistake to locate the norms [of enquiry] at one corner of the triangle – where my peers are – rather than seeing them hovering

out any more comprehensive whole comprised of these. For a totality cannot be partial; a world cannot be a piece of something bigger... So if there is any world, there are many, and if many, none. (McCormick (1996), p.153))

Indeed, we can see that the blurring of the distinction between worlds and true versions, which I have been resisting, is a particular instance of the general phenomenon. Sometimes we talk of worlds and their true versions and sometimes just the versions without the world; contradictory, but according to Goodman, equally right ways to look at the world. So, if Goodman can make good his claim that there are jointly true but contradictory statements, then we would seem to have an argument for worldmaking. And as a corollary, an argument that would undermine physicalism: for if there are many or no worlds, then no sense can be given to the privileged role physics is supposed to play in ontology.

However, the example Goodman offers here does not look promising. The first response will be that in absolute terms neither of these statements, (A1) or (A2) is true, each one is true given some relativisation. So we might say the Earth is stationary *relative to* a geocentric perspective and the Earth moves *relative to* a heliocentric perspective. Goodman, of course, is fully aware of the intuitive appeal of this move (see McCormick (1996), p.154-5). In fact, he wants to tempt his opponent into saying something very much like this. If one does make this kind of perspectival move, Goodman can point out that the talk has shifted from objects (in this case the Earth) to versions; and we can see even more clearly now

over the whole process of triangulation. ... It is not that my peers have more to do with my obligation to say that snow is white than snow does, or than I do." (p.376)

that these are statements referring to different worlds (that is by Goodman's lights different true versions).

Goodman, of course, is right that relativisation manoeuvres such as these involve a change of subject from things to versions of things. But this would only be troubling if there is supposed to be some problem in talking about two versions of the same thing. If, in other words, we could not make sense of the initial problem Goodman confronts us with. The physicalist will no doubt feel he has a good answer here which will avoid irrealism. Just as physicalism provides a background against which we can understand promiscuous realism, it can provide the background against which to make sense of horizontal pluralism. So the physicalist might, for example, invoke a general relativistic space-time understanding of the relation between the Earth and the Sun or they might say that really the Earth goes round the Sun but from the perspective of the Earth, it can appear as though the sun moves. Given such an objective framework on which to cast these two different perspectives, we have no difficulty accommodating the idea that there are two different perspectives on the one world. Just as when the fact that *I see* a clock tower as taller than a building *and you from a different perspective* see the building as taller than the clock tower, this creates no deep philosophical problem. These two perspectives can unproblematically be reconciled against the common background of the actual heights of the building and how one perceives objects from differing perspectives.

Physicalists therefore have a straightforward line of argument to rebut Goodman. First, reject his attack on the version-independence of truth as simply question-begging. Second, insist that the idea of something extra-linguistic, the world, is presupposed in our very use of language. Last, the physicalist is free to

respond to Goodman's challenge of two true but contradictory sentences by either denying the truth of one or other of the sentences or by providing some background against which both sentences can be seen in a sense to be non-conflicting *perspectives* on reality.

The first two of these points, I suggest, should be accepted by everyone – there is something fundamentally wrong with Goodman's idea of worldmaking. However, the third part of Goodman's argument, is a different matter. For those of us pursuing a pluralist alternative, it should stand out as the most interesting. As Goodman himself points out, it is what distinguishes his view from various species of idealism. The claim that the world admits of a plurality of descriptions, horizontal pluralism, is quite separate from Goodman's claims about worldmaking. Can one make sense of this position and avoid irrealism? Or is the choice between denying horizontal pluralism (and thus accepting perhaps physicalism) or embracing irrealism? I have argued at length that physicalism is not credible. So if one rejects physicalism as providing the background against which to make all true statements consistent, then how should one respond to Goodman's third point? Am I stuck between the devil and the deep blue sea, irrealism and physicalism (or some equally implausible patchwork view)? I do not think so. When given an example of two apparently conflicting versions, I think I can also say, without relying on some global metaphysical background picture that certain sentences represent perspectives on the same thing; and translate what that means by giving a more general perspective that removes the apparent contradiction. At least, that is how I would respond to Goodman's challenge regarding how to reconcile sentences (A1) and (A2). However, I do not think that this need always be the case. There are plausible examples of horizontal

pluralism, I maintain. One will be discussed in the following section, where, through the work of Hilary Putnam, I shall try to cultivate an alternative to physicalist and irrealist responses.

6.2.1 Putnam's pragmatic realism

Hilary Putnam has defended a variety of views, with a variety of names (internal realism, pragmatic realism, and sometimes just plain realism), with an equal variety of arguments over the last twenty years. The most famous and most discussed of these is undoubtedly Putnam's (1983) model-theoretic argument. The deluge of criticism that has surrounded this argument (note in particular Lewis (1984) and Devitt (1990)) and the continuing evolution of Putnam's thought make it difficult, perhaps impossible to reconstruct satisfactorily his original intentions, let alone what conclusions we should draw from this argument. In any case, the model-theoretic argument seems to have drifted out of favour with Putnam himself. Although he still discusses it, he does not present it as a simple refutation of any view. I doubt any great insight into Putnam's later philosophy can be gained by a simple study of this part of his work. What I wish to concentrate on instead is a much simpler line of reasoning, similar in structure to Goodman's, which provides an argument for what Putnam calls conceptual relativism.

6.2.2 Conceptual relativism

Consider a world with three individuals, x_1 , x_2 , and x_3 . How many objects are there in this world? Well, that depends on what you mean by object. One answer, the obvious answer, is that there are three objects. (After all, the way in which I have introduced the example suggests that is the right answer.) However, if we

were to adopt a notion of object that admitted mereological sums, then we would say there were 7⁸ objects: x_1 , x_2 , x_3 , x_1+x_2 , x_1+x_3 , x_2+x_3 , and $x_1+x_2+x_3$.

The similarities to Goodman's example should be obvious: we have two apparently true but contradictory statements.

(A1) There are three objects

Asserted in the non-mereologists language and

(A2) There are seven objects

Asserted in the mereologists language.

However, the short way out which Goodman's example is susceptible to, to claim that there is one privileged description from which we can see the others as perspectives, is not available here. For what would we say are the basic parts of the world from which the two perspectives spring? If I say both statements are true but one recognises the existence of mereological sums and the other system does not, then I am already committing myself to the existence of mereological sums, not providing a neutral background against which to judge the two statements as perspectival.

Putnam sees his argument as undermining a favourite metaphor of the metaphysical realist: that there is one dough and many cookie-cutters. There is one dough, according to this metaphor, because there is one way in which reality,

⁸ Ignoring the null object.

so to speak, really is. (Alternatively one might say reality has one *preferred* description.) There are many cookie cutters because we often view reality from different perspectives which take different aspects of that one reality as salient. In this case, unlike the Goodman example, we don't have any sense of what the 'dough' is. Any way of talking about the example either presupposes the mereologists or the anti-mereologists way of looking at things.

How should one react to Putnam's example and what morals should one draw from it? Commentators have reacted in two ways: what I would call the linguistic response and the metaphysical response. The linguistic response is to treat what Putnam has shown us as an example of something which is simply uninteresting and self-evident: that meaning, truth and other semantic properties are language relative. (See in particular Blackburn (1993).) That is to say, the two sentences are not really contradictory at all; rather it is just that the words used in each sentence have different meanings.

Consider these two sentences of eschatology:

(B1) The world will end tomorrow.

(B2) The world will end next week.

These two sentences appear to contradict each other but if sentence (B2) is in some language other than English (call it English*) in which most words mean the same but "next week" means tomorrow, then the two sentences can be seen to assert the same thing. Putnam's more recondite example is an example of the same thing. So there is no special problem of conceptual relativity, just a standard,

boring problem about the correct interpretation of a language. But this would be too quick. Putnam claims that the case he outlines above and the case of the uninterpreted language are not analogous. He agrees that the two sentences, (A1) and (A2), are not contradictory; on pain of inconsistency or a Goodman-style move into a universe of many worlds, he would have to. Nevertheless, there are important differences to note between the example above and the case of the mereological sums. With respect to the two sentences B1 and B2, we have a simple translation of one sentence into the other. Sentence B1 *means* the same as sentence B2. But, claims Putnam, we cannot (or at least cannot so straightforwardly) claim that there is or could be a meaning preserving translation from one version to the other in the case of the language of the mereologist and the language of the non-mereologist. To make the point more vivid, Putnam enriches the story told so far by imagining that one of the objects in the non-mereological world is red and the other two are black. Then the following sentences would be true in the mereologist's language but not in the anti-mereologist's language: "There is an object which is partly red and partly black." Nevertheless, there is a clear sense in which this sentence may be equated with one in the anti-mereologist's language. For example, we might say that if this sentence is true for the mereologist it must be true in non-mereologist language that "there exists at least one object which is red and one which is black". But is this a translation, in any straightforward sense? Well, Putnam remarks it does not seem to be in accord with standard translation practice. We would not be happy to say (or at least it would be controversial) that these sentences in fact had the same meaning.⁹ For example, one has to admit that some of the logical primitives

⁹ People might be tempted to dig their heels in and say that the two sentences do *mean* the same. But without a fully-fledged theory of meaning it is not clear what they would be saying. On a

(exists to pick just one) have different extensions in the two languages. There seems to be a profound conceptual gulf between the mereologist and the anti-mereologist which does not exist in the imagined case of English and English*. Still accepting that Putnam's example is not so trivial as the linguistic interpretation suggests, we may still be left pondering exactly its significance to ontology.¹⁰

Let us turn then to the metaphysical interpretation. I associate this view of Putnam with Ernest Sosa and him alone. Unlike the advocates of the linguistic version, Sosa offers a sympathetic and detailed account of Putnam's argument. He tries several formulations before arriving at version he thinks is of interest and significance. He first rejects the linguistic interpretation; and then suggests a metalinguistic version of the same thesis. That is some sentence in the non-mereological language: "There are 3 objects, not 7" translates into the metalanguage as "In the non-mereological language 'There are 3 objects, not 7' is assertible as true". But any such suggestion as Sosa (and indeed Putnam (1987)) note does not look promising. If it is being claimed that ascent to the metalinguistic level will avoid worries about conceptual relativisation then we need to know what's so special about the meta-level. This looks suspiciously like the cookie-cutter metaphor in linguistic clothing. If not, then to follow this suggestion through consistently would require an infinite sequence of such metalinguistic moves to avoid the conceptual relativity that threatens at each

Davidsonian meaning theory approach, for example, they could not mean the same since the meaning of some of the logical primitives would be different (however this does not mean the sentences would be uninterpretable for Davidson). Putnam clearly sees his original paper as a refutation of Davidson but as we shall see later this involves a misinterpretation.

¹⁰ Another boring way to read the Putnam argument is that it shows us there is no absolute sense of number, only a number of certain things. (This is a very old philosophical point: one Berkeley makes against Locke's suggestion that Number is a primary quality.) Those tempted to read the argument this way should concentrate not on sentences involving numbers of objects but rather questions like: "Do mereological sums *really* exist?"

stage. Clearly then the meta-level has nothing to offer in helping us understand Putnam's views.

Sosa eventually settles on the idea that the only way to understand the significance of Putnam's ideas is by removing them from the linguistic context in which they are offered and transforming them into a more obviously metaphysical claim (or rather a claim which has more obvious implications for metaphysics). He does so by shifting the discussion away from the esoterica of mereological sums and focusing our minds on balls of snow.

6.2.3 Sosa's Putnam

Sosa begins with the observation that most things, natural or artificial, are made up of stuff and parts variously arranged. The way the parts are arranged is essential to the status of the object; change the arrangement of the parts and the object may cease to exist. Sosa uses the example of a snowball:

Thus, the existence of a snowball at a time t and a location l requires that there be a round quantity of snow at l and t sufficiently separate from other snow, etc.; and for that snowball to endure through an interval I , it is required that for every division of I into a sequence of subintervals I_1, I_2, \dots , there must be a corresponding sequence of quantities of snow Q_1, Q_2, \dots , related in certain restricted ways. By all this I mean to point to our "criteria of existence and perdurance of snowballs". (Sosa (1993), p.619)

This way of describing snowballs may strike some as a little pedantic but what Sosa wishes to highlight is that all objects have as it were criteria, however vague, that they must fulfil if they are to be considered to continue to exist. Having made this point about snowballs, Sosa introduces a new entity (or a new word for an entity, lest I be considered to beg the question against certain realists): a snowdiscball. It too obviously has criteria for existence and perdurance. A snowdiscball is an object constituted of snow which may vary from being round to being disc shaped (and everything in between). Hence any snowball will also be a snowdiscball. We are forced to consider the entities – snowball and snowdiscball – distinct, since if one flattened the snowball in question one would destroy it *qua* snowball but not *qua* snowdiscball. But then, so the argument goes, since there are an infinite number of different forms between roundness and flatness that snow might take, there must also be an infinity of different entities¹¹ which have criteria of existence and perdurance which correspond to these small changes in shape. So wherever there is a snowball, there is an infinity of other entities all located at the same point in space and time – Sosa calls this the explosion of reality.

Perhaps understandably Sosa thinks that such a result is undesirable. One way to avoid this ontological extravagance is by adopting Putnam's thesis of conceptual relativity – or at least what Sosa thinks is Putnam's thesis of conceptual relativity. On Sosa's understanding of the doctrine what does and does not exist is not an absolute, objective matter of fact; rather it depends on what conceptual scheme you adopt. In one scheme you may have snowballs but no snowdiscballs (presumably this would be our scheme), in another exactly the

¹¹ Sosa does not consider the idea that vagueness of the parameters of existence and perdurance might save one from admitting an infinite number of entities. Perhaps he believes such vagueness

opposite is true. This is not the only way to avoid the “explosion of reality” according to Sosa. We could deny that any of the objects built up from the parts of matter have a real existence. That is to say, within perhaps certain natural limits, we could claim only the constituents are real, and regard talk of the entity which is so constituted as merely a convenient abbreviation, a way of talking about lots of the real constituents together. So Sosa sees the Putnam style argument as presenting all philosophers with a trilemma. *Either* one accepts the explosion of reality (Sosa also refers to this as absolutism) *or* conceptual relativism *or* some form of eliminativist doctrine. Moreover, Sosa suggests that conceptual relativism may be the least of three evils. Absolutism commits one to a vast ontology without any explanation of why we choose to ignore much of “reality” in our day to day discourse (or to put it another way it fails to explain what is so special about the elements of reality we do fix on). Eliminativism on the other hand seems too high a price to pay. It would make much of our talk about the commonplace literally false.

Wherever most philosophical preferences lie with regard to Sosa’s trilemma, I do not believe he presents a position Putnam would be happy with. Sosa’s reconstruction of Putnam’s argument assumes that one can unproblematically identify the stuff that constitutes the world and then having done so interesting questions can be asked about the status of entities constituted by that matter. In fact, Sosa’s conceptual relativism is a precise account of the “one dough and many cookie cutters metaphor” that Putnam takes himself to be criticising. The matter of the world is the dough and the criteria of existence and perdurance are the cookie-cutters. But the whole point of the recondite example

might threaten his idea of a distinction between schemes.

with the mereologists language and the anti-mereologists language is that there is no privileged set of *objects* and that includes the parts of everyday objects. If Putnam's argument works at all, then it shows the illegitimacy of all three of Sosa's options.

6.2.4 Putnam's Putnam

The real concern of Putnam's argument is, I suggest, to offer a third way between Carnapian conventionalism and Quinean naturalism. Putnam in several papers offers different ways to understand the relation between the following two sentences:

- (1) There exists one object which is red and one which is black. (As stated in the non-mereologists language.)
- (2) There is an object which is partly red and partly black. (As stated in the mereological language)

You might say either (and here I quote Putnam (1987)) that:

- (a) The two sentences are mathematically equivalent.
- (b) The two sentences are logically equivalent.
- (c) The two sentences are neither logically nor mathematically equivalent.
- (d) The first sentence is false and the second is true.
- (e) The two sentences are alike in truth-value and meaning.
- (f) The two sentences are alike in truth-value but unlike in meaning.

(g)The second sentence can be used as an abbreviation of the first, but this is really just a useful make-believe.

Putnam claims that his own view is that “there is no fact of the matter as to which of these positions is correct.” But what does such a position amount to? What Putnam wishes to contend is that certain decisions to describe the world in particular ways are *conventional*. However, he wants to give this way of thinking of the conventionality of certain theses (for example what you take the relationship between sentence A1 and A2 to be) a gloss which allows one to avoid the idea that there should be a sharp analytic/synthetic division.

As I have said, the two opposing views which Putnam asks us to steer between are Carnap’s and Quine’s. The exchange between these two great philosophical heavyweights is part of the folklore of modern analytic philosophy. Philosophers have spent and still spend much energy trying to disentangle precisely the issues at stake between the authors.¹² For the purposes of understanding Putnam the following summary of the basic plot line will do. Carnap distinguished between what he called internal and external questions. The distinction is drawn relative to a linguistic framework. For Carnap a framework was a set of logical and analytic principles which defined the terms therein. Within the framework certain empirical questions might be meaningfully raised and evaluated as either true or false; these are the internal questions. However, questions about the framework itself cannot be meaningfully raised and judged true or false. The way you choose to express your theories, what language or logic you use, is a matter of convention, i.e. an external question. For Carnap, whether

¹² See A. George (2000) for one of the most recent reinterpretations.

or not there are mereological sums is such an external question. If you choose the mereologists' language the answer is yes, if you choose the non-mereologists language the answer is no. But whichever language you choose is purely conventional. The question of whether mereological sums *really* exist cannot arise.

Quine famously criticised Carnap's view for resting on an unworkable analytic/synthetic distinction. There is no way, Quine claimed, to parcel up the meaning of individual sentences into those that depend only upon the language and those which depend on the language plus empirical content. All our beliefs are intimately interlinked and no clear and unequivocal sense can be given to the idea that there are some beliefs which are true and unrevisable in virtue of their meanings (that is the choice of language) alone. So to use just one of Quine's famous metaphors, all our beliefs together face the "tribunal of experience" and any one of them may be revised to accommodate unexpected data.

Putnam's position, as described here, sounds much more like Carnap's – some truths are conventional, just a matter of which language you adopt. In what sense, then, does he steer a middle course between Quine and Carnap? So far he just seems to be following in Carnap's wake.

Like almost all late twentieth century philosophers of significance, Putnam has been deeply influenced by Quine. His earlier work indicates¹³ that he accepts Quine's argument that there is no workable analytic/synthetic distinction or philosophically interesting notion of synonymy. The problem for Putnam then is how to make sense of this idea that some truths are conventional without invoking a Carnap like notion of the analytic/synthetic distinction. This is, I think, a matter

¹³ See, Putnam's *Philosophical Papers, Vol.1*, in particular "The Analytic Synthetic Distinction".

of work in progress but some idea of his evolving position can be found in a recent paper, "Rethinking Mathematical Necessity"(1994). Here Putnam develops a line of reasoning which embraces interpretations of Kant, Frege and Wittgenstein, as well as Quine and Carnap. Putnam sees himself, as he puts it, "detranscendentalizing" an intuition shared by Kant, Frege and the early Wittgenstein that illogical thought is not strictly speaking thought at all. That is to say, Putnam's aim in this paper is to recover a sense in which we may deny that the negations of logical or mathematical statements cannot be comprehended by us and to do so in a way free of what he considers the excess metaphysical baggage of the analytic/synthetic distinction.

6.2.5 Sense and conceptual schemes

The argument Putnam offers is deceptively simple.¹⁴ First, to convince of us of the difference between mathematical and physical statements, he simply highlights some commonplace facts. One can imagine circumstances where one might give up the view that charge was quantised or that one might come to believe that the moon is made of Roquefort cheese (though the story told here would have to be pretty far fetched) but under what circumstances might I give up the idea that $2+2 = 4$? There is none that we conceive of according to Putnam. A statement like " $2+2 = 4$ can be revised" is a statement, he claims, we can give no 'sense' to. We have to be careful with this term 'sense'. It is clear that by 'sense' Putnam does not mean meaning, at least as understood in conventional semantics.¹⁵ As James Conant has put it in a useful introduction to Putnam's most

¹⁴ I am going to ignore Putnam's pre-Carnap historical remarks; these are subtle and, in the case of his interpretation of Frege, at least, controversial.

¹⁵ This is just as well since claiming that sentences like the one above are meaningless would run into well-known objections regarding compositionality.

recent collection of essays: “[Putnam] wishes to drive a wedge between a sentence ‘having a meaning in a language’ and our being able to make sense of the sentence as the assertion of a *claim*.” Since we are unable to provide conditions under which we might recognise the revisability or otherwise of the above sentence, we do not understand what *claim* we are being asked to entertain.¹⁶ Using this notion that some expressions are *for us* senseless, Putnam is able to reject Quine’s naturalism (which places all sentences in the total theory on a par) and simultaneously avoid invoking the analytic/synthetic distinction. For unlike Carnap’s position, the invocation of a conceptual scheme does not require a metaphysics of meaning that makes statements like “ $2+2 = 4$ ” unrevisable, *in principle*. Rather, Putnam is drawing our attention to the fact that even if a sentence is grammatically well-formed and only involves words we can unproblematically attribute meaning to, we may be unable to *understand* that a claim is being made by those words.¹⁷ Quine’s statement about the possible revisability of mathematics is an example of this.

As I have suggested, it is important in establishing some distance between himself and Carnap that Putnam maintains what seems unrevisable for us now, may later seem to become revisable; and this may be done without a revision in the meanings of our basic terms. In other words, our language or our conceptual scheme will develop in such a way as to give a sentence which previously seemed senseless, a sense.¹⁸ Putnam provides a playful example from Wittgenstein’s writings to illustrate the idea that what might seem impossible can come to be so.

¹⁶ Putnam has some other discussions of what is meant by the term ‘sense’. See Putnam (1999), pp. 62-3, Putnam (1992), p. 375.

¹⁷ Note that on Putnam’s conception of what it is for a sentence to have a sense, the statement “‘ $2+2 = 4$ ’ is unrevisable’ must be equally senseless.

¹⁸ On Carnap’s most sophisticated account of his view something like this can happen but since it must involve adoption of a new framework, it must involve changes in meaning.

A court lady once fell into disfavour with the king (one can easily imagine how). The King intending to give her a command impossible of fulfilment, told to her to come to the Royal Ball “neither naked nor dressed”. (1994, p.254)

Now, this does indeed sound impossible but the lady “finds a solution” by coming in a fishnet. Once we know the solution we can see the request is not impossible at all.

Concerning such riddles, Wittgenstein says that we are able to give them a sense only after we know the solution; the solution bestows a sense on the riddle-question. This seems right. (ibid.)¹⁹

The solution makes us appreciate that something that appeared impossible (necessarily not the case) actually is possible. And it has done so in a way which brings to light certain elements of our conceptual scheme which we did not previously focus upon. That is to say, once we have the solution in view, we can see how the words ‘dressed’ and ‘naked’ can be used as they are without paradox and without changing the meaning of either word. As Putnam emphasises, the particular terms used in the example are very important. If we changed the term ‘dressed’ for some phrase we might normally consider synonymous (for example ‘not naked’) the solution to the riddle would not work. For Putnam recognising

¹⁹ This is my interpretation of Putnam’s interpretation of Cora Diamond’s interpretation of Wittgenstein. Fourth-hand analysis like this would be of some concern if my aim were an accurate reading of Wittgenstein. However, it is not; my objective is a plausible account of pluralism. Putnam’s Wittgenstein is important in that account, whether or not it is the real Wittgenstein.

such facts is part and parcel of understanding this as a solution to the riddle. Roughly speaking, Putnam considers this riddle example to provide the model of mathematical discovery (making us aware of new possibilities) and mathematical necessity (marking the boundaries of what is and is not possible).

An obvious objection to Putnam's appropriation of the riddle example as a model of mathematical necessity and the unrevisability of the laws of logic, *at present and by our current lights*, is that it sounds too much like psychologism. It appears as though Putnam is advancing the very radical, and almost certainly false thesis, that what is and what is not necessary is a matter of what we can *imagine* to be the case. He thinks he can escape the charge of psychologism by again invoking the idea of a conceptual scheme. Putnam sees the limitations we suffer in not being able to answer the riddle or prove a mathematical theorem as involving more than lack of imagination. New insights require new concepts. Without such additional concepts we cannot, within our conceptual schemes, even begin to consider whether certain statements are revisable or not. For example, within the structure of concepts possessed by speakers in the eighteenth century Euclidean geometry was unrevisable. That is to say, without the alternative geometries of Lobachevski and Reimann there was no sense to the idea that Euclidean geometry might be wrong. Simply because given any of the conceptual schemes of the time it would not be possible to discern circumstances under which Euclidean geometry might be falsified (and thus by the same token confirmed).

A useful way to view this matter²⁰ is to imagine oneself in the eighteenth century and in the position that our theory of the moment is falsified. We have to make a decision as to what part of that theory to change. Perhaps we set up a

²⁰ This is another example Putnam uses in the paper and refers back to an earlier paper he wrote "It ain't necessarily so".

committee of philosophers to investigate what to do next. These philosophers are anachronistically well-versed in Quinean epistemology and know that they may, logically, change any part of their total-theory. We set up sub-groups of the committee each given the task of revising a different belief (or set of beliefs) which will resolve the anomaly. The first group are allocated the task of changing beliefs at the periphery²¹; the second some more theoretical belief and so on and so on proceeding closer and closer to the centre of our belief system. Given enough time and reasonable consultation with others we would expect each sub-committee near the periphery to be able to formulate some hypothesis which accounts for the data: whether it involves denying the stimulus sentence or changing beliefs about the operation of measuring equipment or tweaking the parameters of the laws of physics, etc. But when we come to the sub-committee asked to save the phenomena by changing the geometry what could they have done? Could they have even understood the request? They might start changing some of the meanings of the terms in their geometry so that perhaps the word “point” refers to lines and vice versa. But although this would be a change, it could not be change that would help save the phenomena; the only way they could really change the geometry is by “discovering” non-Euclidean geometry. In other (or better Putnam’s) words, by changing (and in this case enriching) their conceptual scheme.

The argument in summary is something like this. What is and what is not necessary is a matter of what could and could not be falsified *given* some particular *conceptual scheme*. However, conceptual schemes are fluid, dynamic and open to continuous enrichment. One such enrichment is mathematical

²¹ Here I am employing Quine’s well-known spatial metaphor for our web of belief. Beliefs closest to the periphery are most closely connected to sensory experience; beliefs near the centre (logical

discovery and that proceeds something like the riddle example: by constructing various answers (theorems) we “confer sense” on certain possibilities. Necessity is tied to the idea of possible revisability, which is tied to some particular conceptual scheme. Mathematical discovery involves extending and enriching one’s conceptual scheme by making sense of possibilities not previously recognised.

Although mathematics is the main focus of the article, the implications (as some of the examples discussed suggest) are much broader. The two key concepts here, the concepts that help make up the middle ground between Quine and Carnap, are those of the sense of sentence and the idea of a conceptual scheme. Putnam is attempting to draw significance out of notions that on a Carnapian model collapse back into meaning and framework and on Quinean model language and theory. Now, I imagine a devotee of Quinean doctrine might turn her nose up at the two ideas on which Putnam’s argument is based. She might insist that given one cannot make sense of these ideas in terms of a theory of meaning, the only notion of possibility that is tenable is a psychological one; and that is not one which is of any philosophical significance. Putnam responds thus:

But to convince me that it is possible to imagine the falsity of [a logical truth] you would have to put an alternative logic in the field; and *that* seems a fact of methodological significance, if there is such a thing as methodological significance at all. (1994, p.250)

The concepts of 'sense' and 'conceptual scheme' provide us with a way of discussing acknowledged phenomena. Talk of conceptual schemes highlights the way in which it is not merely lack of imagination that can make us fail to see a proof or how to revise some part of the corpus of our knowledge. There are, as Putnam puts it, methodologically significant leaps which must be made in our conceptual repertoire before we can even begin to entertain certain options. The notion of sense provides some insight into how this process may occur without necessarily involving changes in the meaning of the terms used. The Quinean who wants to relegate this all to psychology becomes blind to the significance of these extensions to our conceptual repertoire.

6.2.6 Other worries

So much for balancing the insights of Quine and Carnap. There may be more fundamental worries that those not immersed in the Quine-Carnap discussion might have regarding Putnam's argument. The argument, as presented, invites two extensions: what I call the strong reading and the weak reading. On the strong reading our conceptual schemes *make* a sentence necessary or not. Parallels with Goodman (and to some extent Carnap) here should be obvious. On the weak reading our conceptual schemes make it impossible to *judge* that any given statement *really* is necessary. Nevertheless, there is a fact of the matter about whether a statement is only necessary relative to our conceptual scheme (Putnam calls this quasi-necessary) or whether it really is necessary (that is necessary for all conceptual schemes).

Both are wrong I think. The strong reading is closer to the position that Putnam wishes to maintain but it goes further than is necessary. We do not make

statements necessary by our conceptual schemes. Rather, the notion of the scheme draws attention to the fact learned from the history of science that certain possibilities cannot be genuinely understood without conceptual enrichment and the fact apparent now that no-one has been able to do this for claims about mathematics or logic. In so doing, Putnam does not see himself as presenting a metaphysical alternative to the pictures of Quine and Carnap, but providing a description “of our lives with our language” (1994, p.259).

Putnam is explicit that the weak reading is not intended.

“Whether a given statement ... “could be revised” depends on whether an alternative theory could be constructed and confirmed and all of the crucial terms “theory”, “confirmation” and “acceptable translation manual” have too much indeterminacy to make application of the principle of bivalence convincing. The illusion that there is in all cases a fact of the matter as to whether a statement is ‘necessary or only quasi-necessary’ is the illusion that there is a God’s-eye view from which all epistemic situations can be surveyed and judged; and that indeed is an illusion.” (p. 258)

One might think that this talk of the spurious nature of the God’s-eye view introduces something new into the argument; something that Putnam’s model-theoretic argument might be required to justify. However, I do not think this is so. First of all, the denial of the God’s-eye perspective should be common ground amongst naturalistically inclined philosophers; and it is those philosophers remember who I believe put forward the only viable case for physicalism. Second,

if we rid ourselves of an ontology of meanings, as we do if we reject the analytic/synthetic distinction, and of facts, as I suggested we should in the last chapter, there is no way to even to begin to make sense of the God's-eye view. For without conceiving of the world as already waiting for description in some favoured language, all we can possibly have to go on in these matters is the way we do, as a matter of fact, talk about things in the world. If the reader remains unconvinced, the next chapter will contain further arguments to support the general point.

6.2.7 Horizontal pluralism

How does this talk of mathematical discovery help our understanding of the example we began this section with? The Putnam's land between the trenches of Carnap and Quine invokes the ideas of sense and conceptual scheme to allow us to see that often developments in conceptual resources are required before certain possibilities can even be entertained. If we consider some of the other cases Putnam cites in support of conceptual relativism – whether there are particles or fields, whether there are really points or merely limits – some people will have differing intuitions about which of these is an empirical question or not. What I suggest Putnam is insisting, is that we should take our decision to adopt one way of speaking or another as simply conventional until we can make some *sense* of what it would be for any of these questions to be empirical. Until, in other words, we can conceive of some sort of situation which would render the decision to speak one way or the other an empirically significant question. Without such a possibility that we can make sense of, we should embrace the thought, which we do in practice anyway, that these two ways 'conceptual schemes' are perfectly

acceptable but different ways of speaking about the same thing. And we simply resist the idea that this in any way commits us to an underlying ontology from which we can see these two views as perspectives. In fact, we have no option but to resist this move since we have no neutral way to conceive of the situation.

Moreover, it should be emphasised there is no impediment in Putnam's example to learning to speak in the other scheme. (How could there be? To understand the example we must understand both schemes.) A speaker of the non-mereological language can learn how to identify what an object is in the mereologist's language. For example, one could point out that the number of objects for the mereologist is equal to the number of nonempty sets of individuals.²² The idea of sense, as opposed to meaning, plays a role here too. If we learn to speak as the mereologist does, then we learn a new sense, not a new meaning of the term object. Putnam's analysis of conventionality allows us to embrace horizontal pluralism without committing us to either some underlying true ontology from which such examples are merely perspectives or Goodmanian irrealism.

In the next section I shall examine two famous papers by Donald Davidson, "Mental Events" and "On the Very Idea of a Conceptual Scheme". Through this discussion, I hope to show how Putnam's considerations in favour of some sort of *pragmatic* horizontal pluralism can be turned into an argument for *pragmatic* vertical pluralism.

²² See Putnam (1994), pp.308-9 for some further discussion.

Chapter 7

Anomalous Pluralism

The following discussion of Davidson is the beginning of my attempt to draw together the elements of a version of pragmatic pluralism. My strategy will be to indicate some tensions between my reading of two of Davidson's most famous papers, "Mental Events" and "On the Very Idea of a Conceptual Scheme". Interpretational fidelity is, I blush to confess it, not my main concern here. My interest in Davidson's Anomalous Monism is to outline a position that escapes the concerns raised about non-reductive physicalism in chapter 2. Such a reading is possible but it comes into conflict with Davidson's rejection of the scheme-content distinction. Drawing on the work of Putnam, I shall argue that this tension can be resolved by adopting a form of pluralism. In fact, I (perhaps immodestly) hope to show that pluralism becomes inevitable.

7.1 Anomalous monism

The position Davidson calls anomalous monism in his paper "Mental Events" arises from three premises: (1) That the mental is causally related to the physical; (2) that singular causal relations are backed by strict laws; (3) there are no strict psycho-physical laws. (Davidson 1993) The first premise is the least controversial: any non-eliminativist, non-epiphenomenal account of the mental will say as much. Premise two has been discussed and criticised at length in my treatment of the completeness of physics and the views of Nancy Cartwright. I shall not elaborate on that here. Reductively minded philosophers will no doubt be dismayed by premise three, whereas for me this is just a particular instance of the fact that there are no strict laws at all. At least initially, I am not concerned with

the truth of any particular premise. It is becoming clearer about Davidson's ontology that interests me first.

How can one make these three premises consistent? Standard accounts of the supervenience relations fall into familiar troubles. To summarise some of the points made in chapter 2: if mental properties only weakly or globally supervene on physical properties, then it becomes difficult to see how mental properties can make any difference to the world. So premise (1) is violated. If we adopt strong supervenience that implies counterfactually supporting relations between the mental and the physical; in other words, psycho-physical laws, hence premise (3) seems to be violated.

A quick and airy dismissal of the notion of the supervenience as means of explicating Davidson's views might seem strange. After all, it was Davidson who was responsible for introducing this term into discussions of the mind-body problem. Nevertheless, I claim, the concept as generally discussed now has little to do with Davidson's original intentions. The most important difference to note between Davidson and the majority of contemporary advocates of non-reductive physicalism is that he defines supervenience as a relation between *predicates*.

A predicate p is supervenient on a set of predicates S if and only if p does not distinguish any entities that cannot be distinguished by S.

(Davidson 1993, p.4)

This shift from talk of properties to talk of predicates changes the way in which we view the causal relation. An assumption made by Kim and many others is that one event causes another *in virtue* of the properties it instantiates. (Again see

chapter 2.) This is categorically rejected by Davidson. The causal relation is, he claims, extensional: if X causes Y, it does so independently of how it is described. No doubt, everyone will accept this up to a point. There is nothing special about any description if all we want to do is pick out the events. Nevertheless, so the thought might go, if we want to understand *why* one event causes another, then we need to know the properties in virtue of which that relation is causal; and these are the properties which figure in laws. Davidson rejects this metaphysical picture because he repudiates the idea that there are intensional entities (properties) which are picked out by certain predicates of a natural language. One event cannot cause another in virtue of instantiating some property because the idea of the world-by-itself (that is without description) having some metaphysically basic or preferred set of properties is itself suspect.¹ Many are troubled by this observation. If it is not the properties of objects that ground the causal relation, then what makes a cause, a cause, they ask? (See for example Crane (1995).) Moreover, the very fact that physical descriptions are special in the sense that it is only in the physical language, according to Davidson, that we can formulate strict laws, might lead us to think that they must pick out some privileged set of properties. After all, it is this special fact about the physical language which makes Anomalous Monism a version of physicalism, right? Whatever the intuitive appeal of this picture is, it is not the way Davidson sees things. There is no special set of properties which glue together events that cause and are caused by other events; if something causes something else then there is no deep, underlying metaphysical reason which explains why this is so.

¹ Davidson suggests that any attempt to explicate the causal relation in terms of properties will violate his insistence that the relation is extensional. I think McLaughlin (1993) has shown that this is not necessarily so.

Once it is clear that Davidson rejects the idea that there is a metaphysical basis to causation, his position becomes clearer. Briefly, it goes something like this: there are these things in the world called events; events can be described using a physical-language or a mental-language; some events are causally related to one another; the events that are causally related to one another are tokens of types of event classified by the physical-language which can be formulated as strict (i.e. exceptionless) laws. The relation between the physical-language and the mental-language is anomalous because there is no systematic way of translating from one to the other. There is no threat of epiphenomenalism because the only thing doing the causing is the events – whichever way you describe them.²

I believe this position is internally consistent. Nevertheless, I think it conflicts with another famous doctrine of Davidson's – his rejection of conceptual schemes.

7.2 Conceptual scheming

"On the Very Idea of a Conceptual Scheme" (hereafter OVICS) rejects what Davidson calls the third dogma of empiricism, the scheme-content dichotomy. Davidson aims his attack at least two types of philosopher. First, there are thinkers like Kuhn and Feyerabend who have claimed different world-views are 'incommensurable'. A typical example given by these writers is to contrast the Ancient Greek astronomer's view of the world with the way a modern astronomer sees it. The Greek thinks of the Sun as a planet but the Earth not; the modern that the Earth is a planet and the Sun a star. There is according to Kuhn no means of drawing a comparison between these two conceptual schemes: the change of

² One might call this *event neutralism*; Davidson occasionally explicitly distances himself from physicalism, suggesting instead that his views are more similar to those of Spinoza and Russell.

meaning in the terms 'star', 'planet', etc. makes this impossible. Hence, the schemes are incommensurable. The other target of Davidson is Quine. In "Ontological Relativity" Quine famously argued that there were many different and incompatible conceptual schemes, each of which could organise our experience, make predictions, etc. in an equally satisfactory way. Since experience is the only external constraint on theory building, there is no fact of the matter, according to Quine, which of these schemes is correct.

Davidson's argument against schemers can be summarised as follows. Conceptual schemes are to be identified with a language (or sets of intertranslatable languages). The criterion for difference between conceptual schemes is failure of translation between the languages. This criterion cannot be met because:

[N]othing, it may be said, could count as evidence that some form of activity could not be interpreted in our language that was not at the same time evidence that that form of activity was not speech behaviour. (Davidson 1984, p. 185.)

Stated baldly like this, Davidson may seem to be conflating the epistemic question of whether we could tell such a language was spoken with the metaphysical question of whether, in fact, there was such a language. In the following, I shall try to indicate why this should be viewed as the conclusion to Davidson's argument, not a self-evident premise. First, we need to appreciate the kind of picture that an advocate of conceptual relativism is offering – Davidson suggests

(See, for example, Davidson (1994).)

two. A conceptual scheme is a way of systematising or organising the *world*; the targets here are Kuhn and Feyerabend. Alternatively, a conceptual scheme is a way of organising the stream of *experience* (and so different conceptual schemes are different ways of so organising). Clearly Davidson has Quine in mind here.

Let us consider the first possibility that what is to be organised by a conceptual scheme is the world. Davidson puts it like this:

It is essential to this idea that there be something neutral and common that lies outside all schemes. This common something cannot, of course, be the *subject matter* of contrasting languages, or translation would be possible...the neutral content waiting to be organised is supplied by nature. (1984, p.190)

If what the neutral content amounts to is the world, then the metaphor of conceptual schemes organizing or fitting is empty. The world being as it is one thing does not lend itself to organisation. If we turn from the world itself to its contents, we shall not do any better. Disagreements and differences regarding organisation of objects must take place against a common, agreed background. (As I emphasised in my discussion of Dupré.) There is therefore no sense that can be given to the noumenal world waiting to be conceptualised. The only world it makes sense to talk about organising is the familiar world, full of familiar objects that we encounter everyday.³

³ Note this is not to deny that there is a world independent of our descriptions as Goodman sometimes wishes to do, but merely to reject the idea of the world-in-itself can illuminate the scheme-content distinction. Davidson's final sentence in OVICS makes this clear. "In giving up the distinction between scheme and world, we do not give up the world but re-establish unmediated touch with the familiar objects whose antics make our theories true or false" (1984, p.192)

The other possibility, that what is organised by conceptual schemes are my unsullied experiences, faces similar difficulties:

The notion of organizing applies only to pluralities. But whatever plurality we take experience to consist in – events like losing a button or stubbing a toe, having a sensation of warmth or hearing an oboe – we will have to individuate according to familiar principles. A language that organizes *such* entities must be very like our own.
(1984, p. 192)

And so translatable into our own scheme. Even if we shift the metaphor of organizing to that of fitting or coping with sensory experience, then Davidson argues the idea of a conceptual scheme fares no better.⁴ If we think of our scheme as coping with the world, then we must be thinking of such schemes in terms of sentences or beliefs expressible in sentences, since only such sentences can be used to make predictions or confront evidence. As Davidson says: “To speak of sensory experience rather than the evidence, or just the facts, expresses a view about the source or the nature of the evidence, but it does not add a new entity to the universe against which to test conceptual schemes.”(p.193)

If the idea of alternative schemes is to have bite here, Davidson claims, evidence must mean more than what we count as such, here and now. Obviously one can have false theories of the world that fit some of the evidence we have actually obtained. But that only shows we can have false theories that fit some of the facts, not that we can have alternative conceptual schemes in any interesting

⁴ The object of Davidson’s attack here is clearly Quine.

sense. The schemer must have in mind something more powerful than that: he must believe that a scheme is relative to all the evidence there ever could be. No extra evidence one way or the other could possibly arbitrate between such schemes. But what can it mean to say that a theory fits or copes with all the available evidence? “[F]or a theory to fit or face up to the totality of possible sensory evidence is for that theory to be true.” (p.193). So the idea of differing conceptual schemes comes down to the thought that there may be two true (or mostly true, since Davidson will allow some disagreement about the details) but non-translatable languages.

In the original paper, Davidson provides a highly compressed and curious sounding argument to get from the premise that a conceptual scheme is mostly true to the conclusion that it is translatable. Tarski’s Convention T, he claims, makes the idea of truth inseparable from that of translation into a familiar idiom. Briefly, the thought is that Tarski has identified an essential feature of truth; namely that a theory of truth should entail sentences of the form: “snow is white” is true if and only if snow is white. When both the quoted sentence on the left of the biconditional and the used sentence on the right are in the same language, the truth of such sentences is obvious. However, if the quoted sentence is in some other language, then that must be a *translation* of the used English sentence. So our understanding of truth in a foreign tongue is dependent on our understanding of translation.

Since Convention T embodies our best intuition as to how the concept of truth is used, there does not seem to be much hope for a test that a conceptual scheme is radically different from ours if that test depends

on the assumption that we can divorce the notion of truth from that of translation. (1984, p. 192-3)

But this does not seem quite right. As Davidson (1990) notes elsewhere: "Tarski's definitions give us no idea how to apply the concept [true] to a new case, whether the new case is a new language or a newly added word to a language." (1990, p.287). The idea of a rival conceptual scheme might reasonably be considered such a new case. So whatever insight Convention T provides concerning the nature of truth, it is not one that we can with a clear conscience apply here.

We might put this point another way. Davidson's theory of meaning depends on the idea that we take truth as basic. It is only by doing so that we can use Convention T as the basis of a theory of meaning. Convention T is used by Davidson to explicate meaning, not truth. It cannot, as he has often emphasised, do both. Given then that Convention T can not, by Davidson's own admission, exhaust the concept of truth, there is no argument yet against saying a scheme is a mostly true but different world-view since our grasp of true (by Davidson's own lights) as used in that sentence must be independent of Convention T.

The principle of charity demands that we try to make Davidson's views appear more plausible. I suggest that the argument offered in this paragraph is elliptical; Davidson is drawing on more than merely Convention T. In the background is his entire theory of meaning. One can perhaps illustrate the point best by turning our attention to what Davidson calls triangulation.

Interpretation and therefore meaning are only possible for Davidson when two or more people interact in a shared environment (hence the triangle). Simplifying slightly, it is through this three-way causal interaction (and the checks

and balances that each corner of the triangle provides) that propositional thought and, with it, the idea of objectivity arise. Speaker and interpreter through their interaction home in on common distal causes of sensations in their environment and with the emergence of language comes the recognition of the possibility of error in one's beliefs regarding those objects. There is no subjective given coming from experience playing a special role in this story. Talk of sensory evidence is just a kind of talk about beliefs, since evidence implies justification and only another belief can justify a belief. That is why to say that one has a world-view which fits the evidence can be to say no more than that a great many of one's beliefs are true. The process of triangulation which speakers of a language inevitable find themselves in ensures the evidence, or better the perceptual beliefs, will concern (*be about*) entities in the shared environment of speaker and interpreter. Such commonality of subject matter implies translatability.⁵

To repeat: the two key points are firstly that only another belief could justify a belief⁶; a conceptual given cannot play this role since at best it could cause beliefs, not justify them. All talk of fitting the evidence, thus, collapses to the idea of being mostly true. Secondly, interpretation and communication more generally presuppose a shared environment⁷; and mostly true beliefs of a shared environment must be translatable using Convention T into our idiom. So Davidson concludes:

⁵ Of course this kind of picture lays a great deal of weight on a shared environment. As I will contend below, at least locally, if not generally, this cannot be guaranteed. This explains certain conceptual asymmetries.

⁶ There is an extremely sophisticated debate involving John McDowell, Hilary Putnam and Robert Brandom (to name only the heavyweights) concerning whether perceptions can be thought to justify beliefs rather than just cause them. Or to put it another way whether one can revitalise epistemology with a direct theory of perception. Interesting and important as this is, I am going to put it to one side. My main interest in Davidson's scheme-content distinction is the scheme-world distinction. That is to say, I'm concerned with how we conceive the world, not how we perceive it.

In abandoning this search [for a theory neutral background], we abandon the attempt to make sense of the metaphor of a single space within which each scheme has a position and provides a point of view. (1984, p.195)

The metaphysical lessons to be drawn are clear. There is no sense to the idea of something waiting to be organised by concepts, whether it be the world, the facts⁸ or experience. Either these ideas are just empty or come to no more than the thought that most of our beliefs are true. Against such a background of largely true beliefs, local disagreements can, of course, be understood but there is nothing like the exciting relativity dreamt of by Kuhn and Feyerabend.

7.3 Events as bare particulars

The conflict between Anomalous Monism and Davidson's rejection of the scheme-*world* distinction is, I hope, clear. If we think of our mental vocabulary as one conceptual scheme and the physical as another, organising the world-of-events, then Davidson himself seems to have a counterexample to OVICS.

One obvious line of response is that both the mental and physical vocabulary belong to the same language. Davidson (1984, Reply to Solomon) himself has said something similar when this objection has been raised. A language, he claims, that contained only physical predicates would not be an

⁷ The idea of a common environment that I have made explicit is implicit in Davidson's talk of all the possible evidence. I think the argument is more convincing spelt out like this because it avoids difficult questions concerning exactly what talk of all the possible evidence might mean.

⁸ Although Davidson does not say so explicitly in this article the same argument applies to the idea that facts are to be organised by experience. Facts are either true sentences or if conceived as

adequate language. One can see why this is so in light of Davidson's theory of meaning. The radical interpreter must begin by imputing intentional attitudes to the speaker of the foreign tongue and correlating them to his own using Convention T and the principle of charity. Thus to be a radical interpreter one must have some concept of propositional attitudes. To cut off mental predicates would be to lose one corner of the triangle of speaker, interpreter and the world that Davidson's claims is essential to interpretation.

All this may be granted. It may well be the case that we could not have a language at all without both a physical and a mental vocabulary. It may well be that this fact supports the intuition that these vocabularies are part of the same language. Nevertheless, we appear to have all the elements Davidson requires to conclude that we have different conceptual schemes: two non-translatable areas of discourse and a neutral something (namely events) waiting to be organised. The real problem is that Anomalous Monism, as I have described it, makes events seem like "something neutral and common that lies outside all schemes". And, if that is so, his own explanation of the mind-body relation does in some way invoke the scheme-world distinction Davidson is seeking to remove from philosophy.

This worry would be removed given some clear criteria of event individuation; one that allowed us to see how it could be the case that a token mental event is identical to a token physical event. Davidson's initial account (1980, "On the Individuation of Events", pp. 163-180) suggested events should be individuated on the basis of their causes and effects. Since what is and what is not a cause or an effect is a token of a physical type, this might suggest Anomalous Monism is a more robustly physicalist position than my presentation allows. One

entities corresponding to sentences are ruled out by Davidson's slingshot argument. See Stephen Neale (1999) for a useful discussion.

might be led to think that the physical vocabulary is doing the work of picking out events. If this is right, then it would be difficult, I think, to resist the conclusion that the physical vocabulary is picking out something special that the mental vocabulary fails to pick out; and familiar charges of epiphenomenalism or downgrading the mental won't be far behind. I don't think this is inevitable. I still believe my interpretation can be maintained; Davidson would merely have to deny that causes *qua* causes are either physical or mental. Anyway, the matter is largely academic since the proposed individuation is obviously faulty.

As Quine (1986) pointed out the criterion is circular “not as a definition but as an individuation”⁹ since causes and effects are themselves events. That is to say, in order to individuate the event we must first individuate the causes and effects of the event. But since the causes and effects are events too, they too must be individuated in terms of their causes and effects (which will include our original event). Davidson has accepted this point. What Davidson considers a suitable replacement, is far from clear. Quine has suggested identifying events with space-time slices. But Davidson seems wary of this proposal, fearing that criterion by itself would lead to the identification of objects with events. In any case, such a suggestion has equally serious drawbacks. If all there is to an event is a slice of space-time that will make many events that we intuitively regard as different, identical. (See the discussion of the Macdonalds in chapter 2 for a more detailed discussion of this claim.) Moreover, the criterion seems at once too lenient and too stringent. If we concern ourselves with only classical physics, then no interesting ontological conclusions could follow from such a view. Providing one is not a dualist, it will just be trivially true that mental events are identical

⁹ A much more detailed account of the problems of token identity is given in Hornsby (1981).

with some physical event (or set of physical events); a host of clearly non-physicalist views, in particular various forms of emergentism would come out as versions of anomalous monism if Davidson followed Quine. That is to say, if you suppose that non-physical mental events can occupy space and they may occupy the same space as their physical bases, as some emergentists clearly did, then you are by Quine's criterion committed to event identity. But intuitively that is wrong. The relation between mental events and physical events the emergentist is trying to articulate is different from that of physicalists or monists like Davidson. On the other hand, once we move to the realm of quantum physics, then the criterion breaks down since quantum states can be superpositions of space and time. Quantum states and events, what one might regard as the best candidates for fundamental physical events, thus do not fit this picture. So, rather embarrassingly for Quine, it seems his criterion of identity cannot be adequate, even at the level of physics.¹⁰

An alternative picture is suggested in a recent article by J. Van Brakel (1999). He interprets Davidson's argument for Anomalous Monism as giving priority to our ordinary talk or what he calls, following Wilfred Sellars, the manifest image. Van Brakel's view is controversial but there are passages in Davidson's work, particularly those that seek to establish some distance from Quine¹¹, that make this a more plausible suggestion than one might think at first blush. Whatever the interpretational rights and wrongs, it is clearly not a suggestion that can help with the individuation problem. There is just no criterion of individuation in ordinary discourse, other than the loose way in which we do as a matter of fact demarcate events. That is to say, we know what event stubbing

¹⁰ These issues are discussed in greater depth in chapters 2 and 4.

our toe or hearing an oboe is; at least we English speakers do. But that provides no general criterion for event individuation from which we can make sense of the idea that some token physical events are identical with some token mental events. Once the idea that there is a neutral stuff onto which the labels of the physical and mental can be applied is undermined, once in other words we drop the scheme-world-of-events distinction, anomalous monism crumbles.

There are obviously two ways to go here. Either one may start to call into question the arguments of OVICS or the premises which make up the argument for Anomalous Monism. Against the background of this work, it is not difficult to guess where my preferences lie.

OVICS complements much of the material discussed in the last chapter. Davidson's arguments against conceptual schemes should be read as further criticism of the one dough and many cookie cutters metaphor that Putnam rejects. It provides conclusive arguments to reject the physicalist idea that we have mooted in our discussions of Goodman and Dupré: viz. the underlying ontology which makes sense of the various ways of describing the world is physicalism. Since physicalists cannot tell us what this physical background from which all other views are merely perspectives is, this view is as empty as the invocation of the world as neutral background on which schemes are cast. When "we abandon the attempt to make sense of the metaphor of a single space within which each scheme has a position and provides a point of view" (p.195), we also abandon this physicalist and metaphysicalist idea of a privileged, though unarticulated role for physics.

¹¹ See his "Reply to Quine", for example, where Davidson claims his main concern is with ordinary language to which science is merely an "exotic suburb".

Davidson, then provides the necessary supplement to Putnam's argument; he demonstrates the emptiness of the God's eye view. That said, there is the embarrassingly obvious problem that the two philosophers seem diametrically opposed on the question of conceptual schemes. Davidson rejects them, Putnam embraces them; and this might make any synthesis of their views appear impossible. The apparent dispute, however, is just that, apparent. Davidson and Putnam simply understand different things by the term 'conceptual scheme'.

To return to the dough and cookie-cutters again: it is tempting to read Putnam as rejecting the dough while retaining the cookie-cutters and Davidson as allowing only one set of cookie-cutters. But that would be to misunderstand both philosophers. The differences, such as they are, are matters of emphasis, not principle. Davidson is drawing our attention to the shared and common aspects of language and belief that we must have in order to understand one another. Putnam is highlighting the way language develops and is enriched by increased conceptual resources. Davidson is right I think to insist that relativism is the wrong term to use for the kind of phenomenon Putnam highlights. Talk of relativism immediately begs the question: relative to what? The answers, as OVICS shows, are either platitudinous or pernicious. But what Putnam calls conceptual relativism, and I call horizontal pluralism, explicitly denies that we should say that differing schemes offer any relativity other than to the scheme (or the language) itself. Putnam's examples highlight the conceptual diversity that exists *within* commonly understood languages such as English. There is no reason why Putnam should not admit that despite their conceptual asymmetry, the mereologist and the non-mereologist share many beliefs – and that provides a background against which we (who can speak either way) can understand their differences. For

example, it is because they share beliefs about the particulate objects, the colour predicates and numbers that they may recognise that they use the concept 'object' with a different sense. Similarly there is no reason why Davidson should object to the kinds of conceptual differences (or asymmetries) existing in one language that Putnam highlights. In fact, his own discussion of the Quinean phenomenon of inscrutability of reference and underdetermination of theory by data suggest he should accept exactly the points that Putnam raises. It is true that the kinds of examples Davidson offers of these phenomena tend to be trivial – the choice, for example, of whether to formulate one's physics using Celsius or Fahrenheit. But the particular point he recognises here should transfer over to Putnam's example. In fact, Davidson's theory of meaning demands that he acknowledges this sort of horizontal pluralism. Truth-theoretic accounts of meaning, like Davidson's, do not single out a unique set of reference (or satisfaction) axioms for singular terms. Nevertheless, some notion of reference is ineliminable from these theories. Hence, the theory must be interpreted as talking about some objects or other. In other words, once we know what sense we are using the term 'object', there can be no difficulty in establishing the truth conditions of our utterances. This is, I take it, just Putnam's point.¹²

All the pieces are in place now to construct a pragmatic pluralism: Putnam's ideas regarding conceptual schemes and sense and the rejection of the privileged perspective from which other views might be seen as relative, what we

¹² Stephen Neale (1999) provides a useful discussion of the significance of these points and their relation to Davidson's rejection of conceptual schemes. There are some subtler issues regarding the details of Davidson's theory of meaning that I am ignoring here which might suggest that the gap between the two philosophers is greater. In particular, Davidson thinks that the requirement that a theory of meaning must be formalisable in terms of a Tarskian theory of truth places, he claims, constraints, on the kind of logic that might be used. Such disagreements though are minor. They may be glossed as concerns about how holistic a theory of meaning should or can be and how central the Tarski-style formal axiomatisation is to a theory of meaning.

have called, following Putnam, the God's eye perspective. I shall try to draw all this together by again returning to anomalous monism.

7.4 Pluralism without global metaphysics

If events *simpliciter* cannot be individuated in any credible way by Davidson's own lights, then we have to take them in the motley way we find them. Anomalous monism accepts that there is a difference between parts of a language, the mental and the physical parts. In the original paper Davidson takes the difference between the mental and physical vocabularies to be mirrored in the difference between intensional and extensional idioms. However, as he confesses, this is unsatisfactory and liable to make all events come out mental.

Setting up principled divisions between different areas of discourse is, I have suggested in chapter three, difficult; at least if one's aim is to defend completeness. Nevertheless, it is a ubiquitous part of our discussion of language. I therefore suggest that we conceive of language generally, and the language of the sciences more specifically, as split up in to many parts: a physical part, a chemical-part, a biological-language part, etc., each different but perhaps also overlapping.¹³ Each part of language will identify a set of events with which it will explain or seek to explain the occurrence other events. Each is in a minimal sense a different conceptual scheme. By which we simply mean that it is as yet not the case that we can translate (that is reduce) one to the other.

I suggest Putnam's discussion of conceptual schemes and sense provides us with the best way to understand the relation between these schemes. The

¹³ Note I am not suggesting we divide the world along the lines of the school curriculum. When I speak of a conceptual scheme associated with physics, chemistry and biology, this is a short-hand way of picking out the multifarious sub-disciplines of each subject, not to mention the fact that

decision to adopt the mereologist or anti-mereologist languages is conventional because we can give no sense to the idea that there is any kind of empirical question which hangs on the decision; hence horizontal pluralism. Similarly, despite much heroic effort, we cannot make any sense of the idea that the different ways of talking that we have in the different sciences are really ways of talking about the very same thing; hence vertical pluralism. Unless we can enrich our conceptual scheme in some way so as to provide a means of allowing us to understand the circumstances under which we might count a mental event as a physical event, unless in other words we have something very like a reduction, then the idea that the mental is just the physical is senseless. But doesn't this kind of talk threaten to cause the world "to disintegrate into a diptych of mental and physical" (Davidson (1980), xi) or a polytych (if there is such a word) of a thousand sciences? No, I claim. There is no need for any Goodman-style pluralism here. The point is not to drive a wedge between different areas of discourse so it seems impossible to relate one to the other. They are related. Quantum physics plays a role in quantum chemistry; broadly physical events are causally implicated in mental phenomena. I do not wish to deny any of this. Rather I want to place these undeniable facts in better view. So we can see that they do not involve any need to privilege one part of the story over the other and that each element is necessary in providing a full account of what is going on.

Anomalous Monism tries to link all these stories by providing one account of events and causation. I have some sympathy with this view. Davidson is right to emphasise the important role causation plays in all of these stories. But in each case our understanding of event and cause is given a slightly different sense. As

they often overlap and complement one another. The explanation of those complementary relations I take to be one of the worthwhile projects of a philosophy of science.

Putnam has shown us with the term 'object', so with 'events'; both "hav[e] an open and ever extendable family of uses" (1994, p. 301). There is unfortunately no one single, general story to tell about the relation of different conceptual schemes to one another. The causal stories are as complicated and pluralistic as any other.

In the concluding section, I shall try to make this brief sketch into a more compelling and concrete alternative and deal with some obvious objections.

Conclusion

The Point of Pluralism

Never mind mind, matter doesn't matter
Nelson Goodman

8.0 What is the world like?

That's a big question and one might expect a correspondingly long and complicated answer, if any answer can be given at all. And one would be right to expect that kind of answer. That's the only answer we can give with a clear conscience. There are electrons, societies, gibbons, books, pork pies, mitochondrial DNA and, of course, philosophers. What these various things do is not something I expect any single person can or could ever hope to know. The best description of how the world is the sum of human knowledge, and the sum of human knowledge at any moment in time is just of course our best descriptions of the world. Never mind, the tautology, if you want to know what the world is like, look to all things we take to be true of the world.

A good place to start, one which has been assumed in this thesis, is with the sciences. If we take the different vocabularies of the different sciences at face value, then we are just as a matter of fact pluralists, committed to a vast array of things and properties and events.¹ Each discipline remains more or less autonomous from others: chemistry and biology both involve independently motivated theoretical concepts, in the same way that psychology inevitably does. To paraphrase Putnam: this is something which I take to be a description of our lives with our science, rather than a piece of metaphysics. Hence, pragmatic pluralism. No doubt this strikes some as glib, an abrogation of philosophical

responsibility and a triumph of theft over hard, honest, philosophical toil. I disagree. Suggesting that physicalism and the like are senseless is not a conclusion that we can reach quickly. It is only something we can be justified in saying after detailed study of the particular claims of the physicalist. In other words, only after going through the kinds of argument offered up in the first half of the thesis. Physicalism is senseless because of its failure to explain the relation between the imagined tidy world of physics and the messy world of contemporary science. The apparatus employed – whether it be supervenience relations or functionalist stories – is not up to the job. Most explanations in terms of supervenience fail on their own terms. They fail to rule out various non-physicalist ontologies. Even proposals that do succeed in fulfilling physicalist desiderata in the abstract, like Yablo's, do not translate into accounts we can understand in the concrete contexts which they are supposed to illuminate. Similarly, functionalist accounts fail to cope with the complexity and richness of intentional phenomena or, in the case of teleosemantics, hold out false hope of the possibility of reduction. Worse still, physicalists cannot provide an account of the physical that does not come into conflict with physics itself or can only offer one so open-ended as to be utterly vacuous. A point which undercuts physicalists of all persuasions – reductive, non-reductive and eliminativist. It is for these reasons (elaborated at length in the first section) that we must take physicalism to be either straightforwardly false or merely offering promissory notes with little suggestion as to how they might be cashed in.

Employing Putnam's notions of sense and conceptual scheme does not provide us with another argument against physicalism; rather it is way of

¹ This is part of my naturalism. It is similar to the attitude Arthur Fine (1986) advocates as natural in his natural ontological attitude.

rationalising the arguments already provided. It is because physicalist have failed to articulate what physics is in their imagined metaphysics and how it relates to other disciplines that we must take their claims as senseless (or false, if we are supposed to understand by physics the discipline practised by those who call themselves physicists). I must confess it would be nice if the terminological innovations of Putnam were an extra argument. It would be great to have a nifty philosophical tool like the positivists' verification principle that could rule out a whole host of problems without the need for detailed argument. But, alas, I have no such labour-saving device. We cannot dismiss physicalism by just declaring 'metaphysics' and moving on. Nor, I should stress, do I mean to cultivate a post-modern, post-metaphysical, post-argument stance from which the arguments of the first section can be viewed as ironic moves in an ultimately hopeless language game. I am not offering argumentative short-cuts or a kind of intellectual balm. When I say physicalism is senseless, I mean that as a summary of the arguments already provided. It is because of the failures of contemporary physicalists to articulate an adequate account of their doctrine that we are left with the long-winded pragmatic pluralist response as the only possible response (at the moment at least) to the question how the world is. Talk of conceptual schemes provides us with a way to refocus attention on the actual theories of our current sciences.²

No doubt this does not alleviate the feeling that this is a glib and unhelpful response. The task of the metaphysician is thought (by some at least) to involve an

² I find some people still do not get what I'm talking about. A way of putting which is entirely illegitimate given my other beliefs is as follows: We do not have the concepts to express physicalism, only the concepts to express a pluralist position. Given we do not yet have the concepts to express physicalism, attempted expressions are either false or, to use Putnam's term again, senseless. This is illegitimate because it implicitly invokes some kind of God's eye perspective. It makes it seem as though there is something to be expressed which might rightly be called physicalism but we have not developed the concepts to express it yet. I hope that it is clear that on my interpretation of Putnam, this view is equally senseless but I hope that gives some sense to the idea that I am trying to get across.

attempt to go beyond the theories of one's time, to provide a unified, underlying picture. All I am doing, it might be argued, is pointing to the mess of science and the failure of current metaphysical alternatives. It is not even a though I am claiming that there is anything intrinsically wrong with metaphysics, only that the accounts of physicalism so far offered are a failure. Such anodyne and general comments hardly constitute a philosophical alternative. In fact, the following response seems apropos: pragmatists are concerned with what difference things make to practice but my position (by my own admission) makes no difference to the way metaphysics is conducted. Therefore, by my own pragmatist lights pragmatic pluralism is self-defeating.

A troubling conclusion, to say the least, if it were true. But it is not. Pragmatic pluralism does make a difference. While, it is correct to say that it is not an attempt to provide an alternative metaphysics, it should, I claim, make a difference to the way we think about metaphysics and the way we think about continuing to do metaphysics.

To see why, we first need to consider what metaphysics is an attempt to do. In the following I shall concentrate on two themes in metaphysical writing. First, what seems to be the most fundamental aspect of metaphysics, is that it tries to say truly and perhaps completely what the world is like. In other words, it tries to address in the most general terms the question with which I began this work. Second, a point often emphasised by philosophers of science is that metaphysics plays an important methodological role in the development of science. I discussed this aspect of metaphysics briefly in chapter one. There, you may recall, I was interested in metaphysics supposed methodological role in science as an argument for the truth of physicalism. Obviously no such argument can be made good; there

is no necessary connection between the fruitfulness of an idea or picture and its truth. Nevertheless, there remains an interesting issue about how to think of the role metaphysics plays in methodology. Pragmatic pluralism, I claim, can suggest new ways to address both aims of metaphysics; and that is enough to render it by pragmatic standards a significant doctrine.

8.1 Global versus local metaphysics

Pragmatic pluralism, at first blush, seems to remove metaphysics from the job of answering the question how the world is and hands it, wholesale, over to science. But this would be to overstate my claims. As I have already said, I have no argument against metaphysics in general, just many separate arguments against the particular metaphysics on offer. I do, however, think that there are bound to be general problems encountered in any attempt to reconcile metaphysics as traditionally pursued and the naturalism that many contemporary philosophers subscribe to. The fundamental difficulty seems to be that metaphysics sets itself up as the ultimate answer to ultimate questions. It represents in that way the end of inquiry. Science, on the other hand is a continuing process of inquiry; it is an ongoing search for better and better answers to specific questions. Since metaphysics wants to be the last word (even if in highly generic and abstract terms) on all matters, it inevitably comes into conflict with the more modest, inquiry led attitude of science. I suggest metaphysicians can become better naturalists by adopting the more modest attitude of science. They should think of metaphysics itself not as the last word on all matters but part of an on-going attempt to understand the world better. I admit pragmatic pluralism can look unsatisfactory, if we view it like traditional metaphysics as the final solution.

However, it should not be seen as that kind of imperialistic metaphysics but rather as a description of the background against which certain kinds of metaphysical project might flourish.

For example, even if we have some clear grasp of conceptual schemes as relating to various different parts of our science (or experience) of the world, we need to understand better the relation between the conceptual schemes: how one set of breakthroughs in one field can influence another; how to understand what is going on when you think you have two conflicting theories. From a properly naturalistic perspective, these questions can only be fruitfully addressed using the concepts of the theories that we currently possess. Pragmatic pluralism should therefore suggest that we approach such metaphysical problems in a piecemeal fashion rather than attempt to do it all in one go. Pragmatic pluralism should make us sceptical about global metaphysics but it should not stop us doing small-scale local metaphysics.

For an example of the sort of local metaphysical problems that can be addressed I want to return to my brief discussion of quantum chemistry. In chapter one, I showed that although quantum mechanics is employed in investigating the properties of molecules, the way the theory is applied makes ineliminable use of knowledge of the structure of the molecule that comes from classical chemistry. A reasonable question for a pragmatic pluralist is: How are we to understand the relationship between the physics and chemistry used in this particular explanation? Well, some of the views of Nancy Cartwright suggest one way to understand the relationship between quantum mechanics and classical physics. Recall that Cartwright argues that the fundamental laws of physics are abstract; they require to be fitted out by a concrete model. Where there is no model, there is

no way to make sense of the idea of those laws applying. Clearly this tallies well with pragmatic pluralism. It does not deny the truth of physical laws, just that we can make sense of the claim that they are at work everywhere. It also provides us with a way of describing the relationship between the two conceptual schemes in this particular example. We may say with Cartwright that quantum mechanics is an abstract set of laws that are fitted out, given a concrete realisation, by the models of classical chemistry. Cartwright's analysis allows us to see the two domains as complimentary rather than as necessarily in conflict. Hence, we are able to avoid the anxieties created by the overdetermination argument. Once we understand the role of physics and that of classical chemistry properly, the very idea that they offer *competing* causes seems absurd. A piece of local metaphysics like this, grounded in a particular understanding of quantum mechanics and classical chemistry, is exactly the kind of metaphysics that pragmatic pluralism should recommend. At least, if one's aim is to say how the world is. Metaphysics pursued like this blends imperceptibly into scientific disciplines that deal with overlapping conceptual schemes, like quantum chemistry. But that, I take it, is what one should expect of any naturalism.

Let me repeat, however, that I do not wish to be censorious. If other philosophers set their horizons higher and wish to attempt more radical, more complete metaphysical projects, then I have no argument to stop them. I simply wish to suggest that one is more likely to be successful in enhancing our understanding of how we think about the world if we try to do metaphysics on a smaller scale. However that may be a matter of taste as much as anything else. Whether your preference is for the panoramic or the close-up, acknowledging pragmatic pluralism should cultivate a modest attitude to metaphysical projects.

So although Robert Brandom may have higher ambitions than me, his general view of metaphysics is consonant with mine when he says:

[T]he task of this genre of creative, non-fiction writing [i.e. metaphysics] is still understood as the engineering of a vocabulary in which everything can be said... The modest metaphysician aims only to codify the admittedly contingent constellation of vocabularies with which her time (and those that led up to it) happens to present her – to capture her time in thought. She sees her task as that of constructing a vocabulary that will be useful to employ to get various sorts of practical grips on [how] things hang together. (Brandom, 2000, p. 181)

Metaphysics modestly pursued in the way Brandom recommends I have no objection to; but it must be rooted in an understanding of other practices and it should be as attuned to differences and disanalogies as it is to connections and similarities.

8.2 Metaphysics as metaphor

What then of the methodological role of metaphysics? As I argued in chapter one this is sometimes overplayed or misinterpreted by philosophers. Various general methodological maxims, like, for example, that it may be fruitful to search for reductionist explanations, do not need (nor could they probably have) a metaphysical underpinning. Nevertheless, I think, a convincing case has been made by some philosophers that metaphysics does indeed play some role in

scientific development. Karl Popper (pp.159-172) gave a particular clear and persuasive account of this in his postscript to *The Logic of Scientific Discovery*. There, Popper set out, in a compelling potted history of science, ten broadly metaphysical theses, ranging from Parmenides block universe to Einstein's vision of a unified field theory that have shaped and influenced scientific research. These he dubbed metaphysical research programmes. He explored one such programme, the development of the modern idea of matter, in detail and showed how extra-scientific ideas helped shape our modern understanding. Popper's rational reconstructions are convincing. Metaphysical claims (though they tend to be more specific than the claims of physicalism) do clearly play some role in scientific advance. The question is how to understand this role.

It may seem that one is at a disadvantage in accounting for any positive role for metaphysics, if one, like me, thinks that most such claims are either false or without a truth-value. However, as will become apparent, I suspect it is an advantage. The key to understanding the connection between general metaphysical claims and methodology is provided by Davidson's theory of metaphor.³

In opposition to almost all other philosophers who have written on the subject, Davidson has claimed that there is no such thing as metaphorical meaning. "Metaphors mean what the words, in their most literal interpretation mean, and nothing more." (1984, p.245) It is Davidson's central claim that if this were not so then what is special or peculiar about metaphor, what divides it from the literal, would be lost. An attempt to graft onto the meaning of words used in metaphor an additional special meaning is an attempt to render the metaphorical

³ I draw heavily on the elaboration and elucidation of Davidson's theory provided by Richard Rorty. See in particular Rorty (1991), pp.162-172

literal. That is to say, once we think of the metaphorical elliptically or indirectly expressing a certain proposition, then it becomes impossible to distinguish it from it having that *literal* meaning. The case of dead metaphors makes this clear. Saying that someone was a pig, no doubt, once was a surprising and interesting use of English. It functioned as a metaphor, directing us to make comparisons between pigs and the person in question. However, habitual and agreed upon use has killed the metaphor off. One of the meanings of the word pig is now, according to my dictionary, “a greedy, obstinate, sulky or annoying person”. Accounts of metaphor that employ a notion of metaphorical meaning are attempts to assimilate all metaphor to dead metaphor. Living metaphors like, “Juliet is the sun” engage us precisely because we do not feel there is a way to paraphrase the sentence plainly so as to convey everything that the metaphor does. And this is so because the metaphor does not convey any meaning apart from its literal meaning. This is not to belittle metaphor. Davidson makes it quite clear that he heartily agrees with those, like Max Black and others, who have stressed the importance of metaphor in helping us see the world in new ways. His point is simply that the notion of an additional metaphorical meaning is not the right way to understand how metaphors function. Metaphor, he claims, is used to evoke all kinds of *effects* in its audience and it is generally these *non-propositional* effects that critics confuse with a separate and additional meaning. As Davidson says: “what we attempt in paraphrasing a metaphor cannot be to give its meaning, for that lies on the surface; rather we attempt to evoke what the metaphor brings to our attention.”(1984, p.262). Metaphor belongs, as Davidson puts it, to the realm of use, not meaning. It is something we do with words, rather than an attempt to convey a proposition using words. The power of metaphor thus lies not in having

a separate meaning but in presenting us with a sentence which has various kinds of effects on us; by prompting us, as opposed to elliptically telling us, to look at things in a new way.

All we need to add to this account for my purposes is that metaphysics as methodology is a species of metaphorical talk. What is being said is false or without a truth-value but it may inspire people to look at the world in new ways. Metaphysical theories are the smelling salts of inquiry; raising our eyes and heightening our awareness of different ways of thinking.

Such an account is, I contend, both an enlightening presentation of the role metaphysics plays in methodology and a fruitful and interesting way to try to develop metaphysical pictures. Certainly, it seems to me the metaphysical views of Nancy Cartwright fit neatly into this picture. Talk of the patchwork of laws or the dappled world⁴ is transparently metaphorical talk. In the introduction to her latest book, Cartwright makes it clear that talk of the dappled world is not, at least primarily, an attempt to provide a true description of the world. Read that way, as I have argued, it simply fails. It is no clearer than for physicalism what accepting the patchwork view commits one to. Both are senseless. But if we consider this talk as metaphorical such objections can be put to one side as simply inappropriate. Cartwright's metaphysics can be reappraised in terms of its *effects*. Read like that we see a very different picture. It becomes clear her main objective is to change our *attitude* towards the sciences. "The worry is not so much that we will adopt the wrong images of the world, but rather we will use the wrong tools with which to change it." (1999, p.12) The metaphor of the dappled world forces us to look askance at reductionist assumptions we might be making and that

⁴ Cartwright takes the term from the poet Gerard Manley Hopkins.

influence our research; it makes us aware of the messy state of contemporary science; and draws our attention away from the abstract to the particular. So even if it fails as a description of how the world is, it may be successful as metaphor in promoting some of Cartwright's wider aims. Generally speaking, I claim, this is the role metaphysics plays in methodology: prompting others to look at the world in fresh ways and inspiring new research.⁵

This should be liberating for metaphysicians. Instead of grafting away with technical relations between p's and m's, they can put their work on a par with poets; they can hope to inspire others to see the world in new ways. They can also hope to point the conversation in new directions. Like Cartwright they may with a clear conscience try to wean us way from puritanical questions about the rightness of theories and force us to ask other questions. Questions about the point of pursuing various programmes of research, of what we hope to do, rather than just what we think we know.

But, as the voices of my Presbyterian ancestors tell me, let's not get carried away. As well as inspiration, there must be perspiration or better yet argumentation. Given that metaphor does not distinguish itself by special meanings, it always open to being misinterpreted as a literal truth. That is to say, the aim of evoking various kinds of effects can be mistaken for the task of trying to describe how the world is. So metaphysicians must warn against this and direct the conversation towards the kind of changes they hope to evoke with their metaphorical talk. Some metaphysicians may of course prefer to direct themselves to the question of how the world is. In that case, they should try to strangle the life out of their favourite metaphors. With the heavy hand and pedantic manner which

⁵ Given this analysis of metaphysics we should probably avoid talk of metaphysical research programmes and say instead that metaphysics may inspire particular scientific research

philosophy specialises in, metaphysicians (or at least some of them) should try to turn the metaphorical into the literal, to make their claims not only inspirational but to provide an argument for their truth.⁶ Such a process is just a particular instance of giving a sense to a previously senseless claim. Of seeing how a form of words connects to a body of knowledge; of learning what is at stake in making a commitment to the truth of one or other thesis. In other words, part of the ongoing process of inquiry.

8.3 The end of the beginning

So I claim pragmatic pluralism does make a difference, a difference in our attitude to metaphysics. It should make us modest when trying to say how the world is. And it should make us bold when trying to inspire new ways of looking at the world. Metaphysics should be rooted in a detailed knowledge of the scientific theories of the here and now. It should also dare us to be imaginative about how to go on from those theories. Metaphysics after pragmatic pluralism should see itself performing these two separate tasks: the inspirational and the argumentative; the poetic and the prosaic; the methodological and the theoretical.

Why pragmatic pluralism matters, in the end, is because it provides the best representation of what we know about the world now; and that is significant not just because of some pedantic desire to be right but also because it should change the way we think about metaphysics. The hope is that we can avoid sterile disputes and make the subject relevant and interesting to practitioners of other disciplines. I confess that I have not done that here. I have been almost

programmes.

⁶ Conversely, of course, one can think of the task that I have been engaged in – showing that various forms of physicalism are false or senseless – as a demonstration that we should construe these utterances as made by metaphysicians as metaphorical.

exclusively concerned with argumentation, with the negative task of demonstrating the failings of particular metaphysical doctrines. But the hope remains that if we attend more closely to the practice of other disciplines and the wider issues of our culture, then metaphysics can be both more accurate and more interesting. Accepting pragmatic pluralism is just the first step.

Bibliography

- Albert, D.Z. (1992) *Quantum Mechanics and Experience* (Cambridge, Massachusetts: Harvard University Press)
- Auyang, S. (2000) *Analysis of Complex Systems* (Proof)
- Beckermann, Flohr, Kim (eds.) (1992) *Emergence or Reduction* (Berlin/New York: Walter de Gruyter)
- Ben-Yami, H. (1999) "An Argument against Functionalism" *Australasian Journal of Philosophy*, 77: pp.320-24
- Bigelow, J. and Pargetter, R. (1987) "Functions", *Journal of Philosophy*, 84: pp. 181-196
- Blackburn, S. (1993) "Enchanting Views", in Clarke and Hale (1993)
- Block, N (ed.)(1980) *Readings in the Philosophy of Psychology, Vol. 1*(London: Methuen)
- (1986) "Advertisement for a Semantic for a Psychology", *Midwest Studies in Philosophy*, pp. 615-678
 - (1993) "Holism, Hyper-analyticity and Hyper-compositionality", *Mind and Language*, 8: pp. 1-26
- Boghossian, P. (1993) "Does Inferential Role Semantics rest Upon a Mistake", *Mind and Language*, 8: pp.27-40
- Boyd, R. (1980) "Materialism without Reduction: What Physicalism Does Not Entail" in Block (ed.)(1980)
- Brandom, R. (ed.) (2000) *Rorty and his Critics* (Malden, Mass. Oxford : Blackwell)
- Brandom, R (2000) "Vocabularies of Pragmatism: Synthesising Naturalism and Historicism" in Brandom (ed.) (2000)

Burge, T (1979) "Individualism and the Mental", in *Studies in Metaphysics*, French, Vehling and Wellstone (Minneapolis: University of Minnesota Press)

Campbell, K (1990) *Abstract Particulars* (Oxford: Blackwell)

Carruthers, P. (1996) *Language, Thought and Consciousness: Essays in Philosophical Psychology* (Cambridge New York : Cambridge University Press)

Carruthers, P. and Botterill, G. (2000) *Philosophy of Psychology* (Cambridge: Cambridge University Press)

Cartwright, N (1983) *How the Laws of Physics* (Clarendon Press: Oxford)

- (1989) *Nature's capacities and their measurement* (New York: Clarendon-Oxford Press)
- (1991) "Can Wholism Reconcile the Inaccuracy of Theory with the Accuracy of Prediction", *Synthese*, 89: pp.3-13
- (1995) "A Reply to Humphreys, Morrison and Eells", *Philosophy and Phenomenological Research*, 55: pp. 177-87
- (1996) "Fundamentalism versus the Patchwork of Laws" in Papineau (ed.) 1996
- (1997) "Where do Laws of Nature come from?" *Dialectica*, 52: pp.65-78
- (1999) *The Dappled World* (Cambridge: Cambridge University Press)

Causey, R. (1977) *The Unity of Science* (Dordrecht: Reidel)

Causey, R. (1981) "Reduction and Ontological Unification: Reply to McCauley", *Philosophy of Science* 48: pp.228-231

Chalmers, D. (1996) *The Conscious Mind : in Search of a Fundamental Theory*(Oxford: Oxford Univesity Press)

Charles, D. and Lennon, K. (1992) *Reduction, Explanation and Realism*, (Oxford: Oxford University Press).

Churchland, P. M. (1979) *Scientific Realism and the Plasticity of the Understanding* (Cambridge: Cambridge University Press)

Clark, P. and Hale, B. (1994) *Reading Putnam* (Oxford: Blackwell)

Crane, T. (1995) "Mental Causation" *Aristotelian Society*, Supp(69): pp.211-236

Crane, T. and Mellor, D.H. (1990) "There is No Question of Physicalism", *Mind* 99, 185-206 reprinted in Moser and Trout (eds.) (1994)

Davidson, D. (1980) *Essays on Actions and Events* (Oxford: Clarendon)

- (1984) *Inquiries into Truth and Interpretation*(Oxford: Clarendon)
- (1986) "Knowing One's own Mind", *Proceedings of the American Philosophical Association*, 60: pp.441-458
- (1990) "The Structure and Content of Truth", *Journal of Philosophy*, 87: pp. 279-328
- (1993) "Thinking Causes" in Heil and Mele (1993)
- (1994)
- (1995) "Laws and Causes", *Dialectica*, 49: pp.263-279

Dennett, D. (1996) "Cow-Sharks, Magnets and Swampman" *Mind and Language*, 11: pp. 76-7

Devitt (1997) *Realism and Truth*, 3rd edition (Oxford: Blackwell)

Devries, W (1996) "Experience and the Swamp Creature", *Philosophical Studies*, 82: pp.55-80

DeWitt, B. (1970) "Quantum Mechanics and Reality" *Physics Today*, 23: pp.30-35

- Dupre, J. (1996a) *The Disorder of Things*, 3rd edition (Cambridge, MA: Harvard University Press)
- Dupre, J. (1996b) "A reply to Wilson", *British Journal for the Philosophy of Science*, 47: pp. 441-444
- Ehring, D. (1996) "Mental Causation, Determinables and Property Instances", *Nous*, 30: pp.461-480
- Field, H (1978) "Mental Representations", *Erkenntnis*, 15: pp.9-61
- Fine, A (1986) *The Shaky Game* (Chicago: Chicago University Press)
- Fodor, J. (1974) "Special Sciences", *Synthese*, 28: pp.97-115
- Fodor, J. and LePore, E. (1991) *Wholism: A Shopper's Guide* (Oxford Cambridge, MA : Blackwell)
- Goodman, Nelson (1984) *Of Mind and Other Matters* (Cambridge, MA: Harvard University Press)
- Hacking, I. (1983) *Representing and Intervening* (Cambridge: Cambridge University Press)
- Haugeland, J. (1984) "Ontological Supervenience" *Southern Journal of Philosophy* 22, Supplement: pp.1-12
- Heil, J (1992) *The Nature of True Minds* (Cambridge: Cambridge University Press)
- Heil, J (1998) "Deconstructing Supervenience" *The European Journal of Philosophy*
- Heil, J. and Mele, A. (eds.)(1993) *Mental Causation* (Oxford: Oxford University Press)
- Hellman and Thompson (1975) "Physicalism: Ontology, Determination and Reduction", *Journal of Philosophy*, 72: pp.551-564

- Hellman and Thompson (1977) "Physicalist Materialism", *Nous* 11: pp.309-345
- Hempel, C. (1966) *Philosophy of Natural Science* (Englewood Cliffs (N.J.): Prentice Hall)
- (1980) "Comments on Goodman's *Ways of Worldmaking*", *Synthese*, 45: pp.143-9. Reprinted in McCormick (1996)
- Hooker, C.A. (1981) "Towards a General Theory of Reduction", *Dialogue* 20: pp.38-59, 201-306, 496-529.
- Horgan, T. (1993) "From Supervenience to Superdupervenience: Meeting the demands of the Material World", *Mind* 102: pp. 555-586.
- Hornsby, J. (1981) "Which Physical Events are Mental Events?" *Aristotelian Society Proceedings*, 81: pp.73-92
- Hughes, R.I.G. (1989) *The Structure and Interpretation of Quantum Mechanics* (Cambridge, MA: Harvard University Press)
- Jackson, F.(1986) "What Mary didn't Know" *Journal of Philosophy*, 83:pp. 291-295
- (1996) "Mental Causation" *Mind*, 105: pp.377-413
- Kain, S. (2001) *Constructing the Perfect Ground State Potential for Water*, PhD, University of London
- Kemney and Oppenheim (1956) "On Reduction" *Philosophical Studies*, 7: pp.10-19
- Kim. J. (1989) "The Myth of Non-reductive Materialism" reprinted in *Contemporary Materialism*, ed. Moser and Trout (eds.)(1994)
- (1992) ""Downward Causation" in Emergentism and Non-reductive Physicalism" in Beckermann, Flohr, Kim (eds.) (1992)

- (1993) *Supervenience and Mind: Selected Philosophical Essays* (New York, NY, USA : Cambridge University Press)
- (1996) “Non-reductionist Troubles with Downward Causation” in Heil and Mele (eds.)(1996)
- (1998) *Mind in a Physical World* (Cambridge, MA : The MIT Press)
- (1999) “Supervenient Properties and Micro-Based Properties: A Reply to Noordhof” *Aristotelian Society Proceedings*, 99: pp. 115-8

Kripke, S (1970) *Naming and Necessity* (Oxford: Blackwell)

LePore, E and McLaughlin, B (eds.) (1985) *Actions and Events: Perspectives on Philosophy of Donald Davidson.* (Oxford: Blackwell)

LePore, E (ed.) (1986), *Truth and Interpretation: Perspectives on the philosophy of Donald Davidson* (Oxford: Blackwell)

Lewis (1983) “New Work for a Theory of Universals” *Australasian Journal of Philosophy*, 61, pp.343-377

- (1984) “Putnam’s Paradox” *Australasian Journal of Philosophy*, 62: pp. 221-236

Macdonald, G. and Macdonald, C.(eds.) (1995) *Philosophy of Psychology*(Oxford Cambridge, MA: Blackwell)

- (1995) “Supervenient Causation” in the above
- (1995) “How to be Psychologically Relevant” in the above

Manning, R. (1997) “Biological function, selection and reduction”, *British Journal for the Philosophy of Science*, 48, pp. 69-82

McCormick, Peter (ed.) (1996) *Starmaking* (Cambridge, MA London: MIT Press)

McDowell, J. (2000) "Towards Rehabilitating Objectivity" in Brandom (ed.) (2000)

McLaughlin, B. "On Davidson's Response to the Charge of Epiphenomenalism" in Heil and Mele (1993)

- (1997) "Supervenience, Vagueness and Determination", *Philosophical Perspectives 11: Mind, Causation and World*

Mellor, D. H. (1995) *The Facts of Causation* (London: Routledge)

Melynk, A. (1998) "How to Keep the Physical in Physicalism", *Journal of Philosophy* 95: pp.662-637

Merricks, T. (1998) "Against the Doctrine of Microphysical Supervenience" *Mind*, 107: pp.59-71

Millikan, R. (1984) *Language, thought, and other biological categories : new foundations for realism* (Cambridge, MA: MIT Press)

- (1993) *White Queen psychology and other essays for Alice* (Cambridge, MA: MIT Press)

- (1995) "A Bet with Peacocke" in Macdonald and Macdonald (1995)

- (1996) "On Swampkinds", *Mind and Language*, 11: pp.103-117

- (1999) "Wings, Spoons, Pills and Quills: A Pluralist Theory of Function", *Journal of Philosophy*, 96: pp.191-206

Moser, P and Trout, J. D. (1995) *Contemporary Materialism: a reader* (London, New York: Routledge)

Morrison, M. (1995) "Capacities, tendencies and the problem of singular causation" *Philosophy and Phenomenological Research*, 55: pp.163-68

Neale, S. (1999), "On Representing" in *The Philosophy of Donald Davidson*

Neander, K. (1996) "Swampman meets Swampcow" *Mind and Language*, 11: pp. 118-129

Oppenheim, F and Putnam, H. (1958) "The Unity of Science as a Working Hypothesis." *Minnesota Studies in the Philosophy of Science*, Vol. 2

Papineau, D (1987) *Reality and Representation* (Oxford: Blackwell)

- (1993) *Philosophical Naturalism*, (Oxford: Blackwell)
- (ed.) (1996) *The Philosophy of Science* (Oxford: Oxford University Press)
- (1996) "Doubtful Intuitions" *Mind and Language*, 11: pp.130-2
- (1998) "Mind the Gap" *Philosophical Perspectives 12: Language, Mind and Ontology*
- (unpublished) "A Reply to Brandon-Mitchell and Jackson"

Papineau, D. and Spurrett, D. (1998) "A Note on the Completeness of 'Physics'", *Analysis*, 59, pp. 25-29

Peacocke, C. (1995) "Concepts and Norms in a Natural World" in Macdonald and Macdonald (1995)

Poland, J. (1994) *Physicalism* (Oxford: Clarendon Press)

Popper, K. (1982) *Quantum Mechanics and the Schism in Physics* (London: Routledge)

Post, J. (1991) *Metaphysics: A Contemporary Introduction* (New York: Paragon House)

Prades, J. and Corbi, J. (2000) *Minds, Causes, and Mechanisms : a Case Against Physicalism* (Oxford: Blackwell)

Preston, B. (1998) "Why is a Wing like a Spoon? A Pluralist Theory of Function." *Journal of Philosophy*, 95: pp. 215-254

- Price, H. (1992) "Metaphysical Pluralism", *Journal of Philosophy*, 89: pp.387-409
- Putnam, Hilary (1960) "Minds and Machines" in Hook (ed.) *Dimensions of the Mind* (New York: New York University Press)
- (1975) "The meaning of 'meaning'" *Language, Mind and Knowledge, Minnesota Studies in the Philosophy of Science*, vol. 7
 - (1979) *Philosophical Papers, vol 2: Mind, Language and Reality* (Cambridge: Cambridge University Press)
 - (1987) "Truth and Convention" *Dialectica*, 41: pp.69-77
 - (1990) *Realism with a Human Face* (Cambridge, MA: Cambridge University Press)
 - (1992) *Renewing Philosophy*(Cambridge, MA London: Cambridge University Press)
 - (1994) *Words and Life*(Cambridge, MA: Cambridge University Press)
 - (1999) *The Threefold Cord*(New York: Columbia University Press)
- Rorty, R. (1991) *Objectivity, Relativism and Truth* (Cambridge: Cambridge University Press)
- (2000) "Reply to McDowell" in Brandom (ed.) (2000)
- Sahorta, S. (2000) *Genetics and Reductionism* (Cambridge: Cambridge University Press)
- Savellos and Yalçin (ed.) (1995) *Supervenience*, (Cambridge: Cambridge University Press)
- Shoemaker, S. (1984) *Identity, Cause and Mind : Philosophical Essays* (Cambridge: Cambridge University Press)
- Sklar, L. (1995) *Physics and Chance* (Cambridge: Cambridge University Press)

- Smith, P.(1992) “Most Reductions and the Unity of Science” in Charles and Lennon (1992)
- Sober, E. (1999) “Physicalism from a Probabilistic Point of View”, *Philosophical Studies*, 95: pp.135-174. (Page numbers in text refer to web version.)
- Sober, E and Segal, G. (1991) “The Causal Efficacy of Content”, *Philosophical Studies*, pp. 1-30
- Sosa (1993) “Putnam’s Pragmatic Realism”, *Journal of Philosophy*, 95: pp. 605-626
- Sturgeon, S. (1998) “Physicalism and Overdetermination”, *Mind*, 107: pp. 411-432
- Van Brakel, J. (1999) “Supervenience and Anomalous Monism”, *Dialectica*, 53: pp. 3-24
- Waters, K.(1998) “Causal regularities in the Biological world of Contingent Distribution”, *Biology and Philosophy*, 13: pp.5-36
- Wilkerson (1993) “Species, Essences and the Names of Natural Kinds” *Philosophical Quarterly*, 43: pp. 1-19
- Wilson, J (1999) “How Superduper does a Physicalist Supervenience need to be?” *Philosophical Quarterly*, 49: pp.33-52
- Wilson, R (1996) “Promiscuous Realism”, *British Journal for the Philosophy of Science*, 47: pp. 303-316
- Witmer, G. (2000) “Locating the Overdetermination Argument”, *British Journal for the Philosophy of Science*, 51: pp. 273-286
- Worrall, J. (1989) “Structural Realism: the Best of Both Worlds”, *Dialectica*, 43: pp.99-124. Reprinted in Papineau (ed.) (1996)
- Yablo, S. (1992) “Mental Causation” *Philosophical Review*, 101: pp. 245-280

- (1997) "Wide Causation" *Philosophical Perspectives 11: Mind, Causation and World*