The Super-Overdetermination Problem

John Donaldson

MA, MSc

Submitted in fulfilment of the requirements for the Degree of PhD

School of Humanities

College of Arts

University of Glasgow

October 2014

For
My family

Abstract

I examine the debate between reductive and non-reductive physicalists, and conclude that if we are to be physicalists, then we should be reductive physicalists. I assess how both reductionists and non-reductionists try to solve the mind-body problem and the problem of mental causation. I focus on the problem of mental causation as it is supposed to be faced by non-reductionism: the so-called overdetermination problem. I argue that the traditional articulation of that problem is significantly flawed, and I show how to articulate it properly: what I call the 'super-overdetermination problem'. In doing so, I demonstrate that the problem of mental causation faced by non-reductionism is in fact a special case of the mind-body problem, as faced by non-reductionism, and that the former can't be solved independently of the latter. I then assess the prospects for a particular family of non-reductive views that I call immanentism, and show that they fail to solve the super-overdetermination problem. Finally, I put forward two arguments to support the conclusion that physicalism entails reductionism. Both arguments establish, via distinct reasoning, the proposition that mental property instances are identical to physical property instances; and then each argument employs the inference, which I also defend, that if mental instances are physical instances, then mental properties are physical properties; hence, reductionism follows.

Contents

Acknowledgements

I am deeply indebted to my family for their support throughout my studies, but in particular during the completion of my PhD. My father, William, my brother, Adam, my sister, Janet, and my late mother, Loraine, supported me in so many ways it would be impossible to list them all here. Without their backing, this thesis would not have been completed. I owe a particular debt of gratitude to my father, who, either through specific instruction or simply by setting a good example, has taught me every day, from my very first day until today, how to be a scholar, a parent, a person. My life has also been filled with the joy of being the father to an irrepressible five-year-old, Lucia, who has been an unending source of light for me during many grey days.

I am in the debt of many friends and colleagues with whom I have shared the journey towards PhD submission. But in particular, Robert Cowan has been an invaluable source of personal and philosophical advice over many years, and a very good friend. I am grateful to my friend and colleague Umut Baysan, with whom I share a number of philosophical interests, for numerous highly illuminating discussions of many of the issues in this thesis, in particular in relation to the subset view on which Umut shared much of his specialist knowledge with me.

I wish to thank Robert, Umut, Neil McDonnell, Andrew MacGregor and Stephanie Rennick for reading parts of my thesis and giving me their valuable thoughts. And I am particularly grateful to my father who proofed my thesis in its entirety.

I feel exceptionally fortunate to have benefited from supervision of the highest standard. I am the son of an excellent teacher, so I know high quality instruction when I see it, and the supervision I received was a model of good pedagogical practice. Not just in what was taught, but the manner in which it was delivered. Dr David Bain supervised from me from the very beginning until the end of my thesis, and guided me with great skill through many philosophically choppy waters. Dr Stephan Leuenberger jointly supervised me along with Dr Bain during the second half of my thesis, and I found his expertise in the metaphysics of mind to be particularly invaluable. During the first half of my thesis, Professor Fiona Macpherson was my second supervisor, and provided very valuable instruction and support during that time.

When I think of the kind of philosopher and teacher I would like to be, I think of the philosophers who taught me.

I also feel exceptionally fortunate to have undertaken my PhD in the philosophy department at the University of Glasgow. The intellectually stimulating and collegial environment which exists in the department is, I believe, quite exceptional, and I am privileged to have spent my time as a novice scholar within the confines of that environment. I wish to acknowledge the contribution of every member of the faculty and graduate community to the creation of that extremely hospitable environment.

I would also like to take this opportunity to mention the influence of two philosophers whose contributions to the debate shaped my own thinking in profound ways: Professor Jaegwon Kim and Professor Stephen Yablo. Although I disagree strongly with both Professor Kim and Professor Yablo in my thesis, the influence of their work on my thinking and philosophical development cannot be overstated. I am also very grateful to Ruth Perry, my stepmother, who, with kindness and foresight, introduced me to Professor Yablo. It was following that meeting, and becoming immersed in Professor Yablo's work, that the development of the core ideas which form the heart of my original contributions in this thesis took place. And I am also grateful to Professor Yablo for agreeing to read a large part of my thesis, and for his supportive and helpful comments upon doing so.

Finally, I would like to thank a number of institutions for extending financial support to me during my studies. The Sir Richard Stapley Educational trust and the McGlashan Trust gave me grants over a number of years that were immensely helpful in pursuit of my PhD. I am also grateful to Professor Macpherson who, as Director of the Centre for the Study of Perceptual Experience, extended financial support to me at the beginning of my studies.

I could easily have written acknowledgements that ran to several pages, and yet still have failed to recognise fully all those who have helped me in one form or another. I am very fortunate to have met so many kind hearts and friendly faces along the way.

What consciousness is, we know not; and how it is that anything so remarkable as a state of consciousness comes about as the result of irritating nervous tissue is just as unaccountable as the appearance of the Djinn when Aladdin rubbed his lamp.
-T. H. Huxley[1]

Materialism's Problems of Consciousness

There are supposed to be two major problems of consciousness. The problem of explaining how consciousness can arise out of physical processes, and the problem of explaining how consciousness can make a causal difference in a world where effects already have physical causes. The djinn metaphor captures the first problem, and Huxley had another metaphor to capture the second:

> The consciousness of [animals] would appear to be related to the mechanism of their body simply as a collateral product of its working, and to be as completely without any power of modifying that working as the steam-whistle which accompanies the work of a locomotive engine is without influence upon its machinery (1874: 236).[2]

Huxley's metaphors illustrate what have come to be known as the mind-body problem and the problem of mental causation, particularly as those problems are faced by philosophers who believe that there is a material world, and that psychological occurrences–mental events–occur within that world in accordance with natural laws.

In Huxley's time, the truth of materialism–the doctrine that in some sense everything is material–was very much an open question, and although that remains the case today, it is to a much lesser degree. Contemporary discussions of the mind-body and mental causation problems most often take place within a presumptive materialist framework, or on the assumption that materialism is true; and all that remains, many think, is the task of convincing the sceptics, or of resolving in-house materialist disputes. It is this latter task

[1] (1866: 193).

[2] Although it is true that here Huxley was concerning himself with the question of whether the mental states of non-human animals ('brutes' as he called them) were epiphenomenal (whether animals were 'automata' as Huxley put it), he was not speciest in his conclusions. As Huxley states, 'to the best of my judgment, the argumentation which applies to brutes holds equally good of men; and, therefore, that all states of consciousness in us, as in them, are immediately caused by molecular changes of the brain-substance. It seems to me that in men, as in brutes, there is no proof that any state of consciousness is the cause of change in the motion of the matter of the organism' (1874: 239).

which will be our focus in these pages. That in-house materialist dispute is between reductive and non-reductive materialists, or 'physicalists' according to the preferred contemporary nomenclature.

With regards to the mind-body problem, reductionists take a hard line: the mental just *is* material in a strict sense; non-reductionists take a softer line: the mental is 'nothing over and above' the material, yet is distinct from it. I defend reductionism by arguing that one can be a physicalist only if one is a reductionist.

With regards to the problem of mental causation, it is widely believed that it is a problem for non-reductionism, but not reductionism. I argue that this belief is false, and based on a widespread failure to grasp what the problem of mental causation as faced by materialism really is. The problem of mental causation, I suggest, is in fact a special case of the mind-body problem, rather than a stand-alone problem in its own right which can be solved independently of solving the mind-body problem.

I complete these arguments in four main parts. First, I examine the genesis of the debate between reductionists and non-reductionists, and characterise the two views. Second, I show how the problem of mental causation as faced by non-reductionism, the so-called 'overdetermination problem', has been misarticulated, and I suggest how to articulate it properly. Third, I assess a new wave of non-reductive views that I call 'immanentist', which are of interest in large part because they seem to promise a solution to what I take to be the proper articulation of the overdetermination problem. I show how immanentism fails to deliver on that promise, and seems to be something of a wrong turn in the development of NRP. Fourth, in the light of immanentism's failure I assess the debate between reductionism and non-reductionism, comparing what I take to be the two strongest versions of each. I then put forward two arguments which show that physicalism requires reductionism.

We will shortly consider the general argument in more detail by summarising the contents of each chapter, but first I wish to make a few brief remarks about terminology.

I said above that I am here resolving an in-house dispute among physicalists. What I do not try to resolve in these pages is the difficult question of what it means, exactly, to describe an entity as 'physical'. I will, however, make the following remarks as a guide to my usage. I treat the terms ‘matter’, ‘material’ and ‘physical’ as synonyms; similarly with ‘mental’, ‘mind’ and ‘psychological’.

I use ‘mental’ to refer to entities such as propositional attitudes, perceptions, emotions, and sensations. At this stage, I leave open, as I must, whether mental entities are physical, or not; thus my use of ‘mental’ is neutral, unless stated otherwise.

I use the term ‘physical’ in three ways:

(i) ‘microphysical’ to refer to that which is designated by the terms essential to complete microphysical theory, which, if current microphysical theory is near enough to the truth, will include entities like quarks, bosons, and the laws that govern them and their various siblings;
(ii) ‘physical’ to refer to that which is designated by the terms essential to complete physical theory more generally, which, if current physical theory is near enough to the truth, will include entities like quarks, bosons, chemical elements, cells, neurons, and the laws that govern them and their numerous siblings;
(iii) ‘broadly physical’ to refer to the entities which either are, or depend for their existence on, physical entities.

Obviously, I offer these remarks not to provide definitions, as such. I simply intend to make clearer what the terms I use refer to out of all the various entities we intuitively identify as mental and physical. As I am adjudicating an in-house dispute between physicalists, a genuine definition of ‘physical’ is, thankfully, not required. In the above, I have used the phrase ‘if current (micro)physical theory is near enough to the truth’. I leave open whether or not current theory is close enough. If it isn’t close enough, then physicalists will lose the wager that future completed physical theory will be a recognisable descendent of current physical theory, such that the former resembles the latter sufficiently for our current talk of ‘physical’ defined in terms of the above to make sense. For what it’s worth, I think the odds for a recognisable physicalism at the end of science seem reasonable enough, but nothing will rest on that claim here. Whether or not the mind-body problem can be stated clearly at all is an interesting question (see Montero 2009; Crane and Mellor 1990; Hempel 1980), but the arguments below are addressed solely towards those who think that the mind-body problem can be the focus of sensible debate, and thus that the terms the problem is framed with–'mental' and 'physical'–do not offend the Gods of clarity and good sense.

Given our understanding of what physical means, we can state what the in-house dispute between reductionists and non-reductionists amounts to: whether the mental is merely broadly physical, as the latter claim, or whether the mental is physical, as the former claim.

In what follows, I will use the acronyms 'RP' and 'NRP' to refer to reductive and non-reductive physicalism, respectively. We turn now to a summary of the contents of each chapter.

Chapter 1: Reductive Versus Non-Reductive Physicalism.

We begin by considering the arguments in favour of a key claim: that reductionism is best understood as the thesis that mental properties are identical to physical properties, and contrast different conceptions of reduction in doing so: ontological versus theoretic. We then examine non-reductionism's founding ideas–multiple realisation and the disunity of science–before moving on to a consideration of Donald Davidson's anomalous monism as a candidate NRP view; we conclude that it is not physicalist. We close the chapter by asking whether a view I call 'disjunctive physicalism' is a non-reductive position as some of its proponents claim. I conclude that it is not; disjunctive physicalism is in fact a version of RP.

Chapter 2: The Super-Overdetermination Problem.

We discuss the original intuition on which is based the thought that NRP faces an overdetermination problem, what I call 'Malcom's intuition' and lay down conditions for correctly articulating that intuition such that a genuine problem for NRP is stated. A genuine problem is then articulated: what I call the 'super-overdetermination problem', which is shown to be a special case of the version of the mind-body problem faced by NRP: the superdupervenience problem. We then contrast the super-overdetermination problem with four disambiguations of the traditional articulation of Malcom's intuition, and show how each fails to articulate a genuine problem for NRP. This is followed by a detailed critique of the most famous articulation of Malcom's intuition: Jaegwon Kim's exclusion and supervenience arguments. Penultimately, we discuss how RP faces what I call the 'epistemic-overdetermination' problem. We finish the chapter by diagnosing the half-century persistence of the flawed articulations of Malcom's intuition.

Chapter 3: Immanentism Assessed.

First we note that there is an interesting group of cutting-edge NRP views which try to solve the traditional overdetermination problem but do so in a manner which seems to offer hope of an NRP solution to the super-overdetermination problem. Each version of immanentism appeals to a different relation between mental and physical causes such that it might seem true to say that mental causes are 'immanent in' physical causes: the determinate/determinable, constituted/constituter, and set/subset relations. Three criteria are laid down with which to assess whether immanentism can solve the super-overdetermination problem. The rest of the chapter is taken up with examining each immanentist view in turn, and showing that each fails to meet the three criteria.

Chapter 4: Physicalism entails Reduction

Given the ground covered in the previous chapters, we consider the state of play with regards to the RP-NRP debate, and consider the best version of reductionism and non-reductionism. Then, two general problems for NRP are considered: what I call the 'instance problem' and the 'ghost problem'. On the basis of these problems, I argue that mental instances must be physical instances, and if all mental instances are physical instances, then mental properties are physical properties and reductionism follows. Thus, I conclude, if we are to be physicalists, then we need to be reductive, rather than non-reductive physicalists.

Chapter 1

Reductive Versus Non-Reductive Physicalism

Chapter Summary

We begin by considering the question of what mental-physical reductionism is: a mental-physical property identity thesis. We contrast theoretic and ontological approaches to reduction, defending the latter as doing greater justice to the concerns at the heart of mental-physical reductionism. Following this, we survey NRP–outlining the core ideas on which the view was founded: multiple realisation and the disunity of science. We also discuss Donald Davidson's anomalous monism–claimed by some to be a variety of NRP–but judged here not to be a physicalist view, and so to fall outwith the bounds of NRP. Finally, we turn to the interesting position of disjunctive physicalism, the proponents of which identify themselves as non-reductionists, despite accepting that mental properties are identical to physical properties. However, it is argued, the reverse is true: disjunctive physicalism is in reality a reductionist position, which, if it works as its proponents advertise, demonstrates the compatibility of reductionism with NRP's founding principles of multiple realisation and the autonomy of the various, dis-unified, sciences.

The word 'reductionism' seems by now to have acquired a negative, faintly disreputable flavour [...] Being a reductionist is a bit like being a logical positivist or member of the Old Left–an aura of doctrinaire naiveté hangs over such a person.
- Jaegwon Kim[3]

1.1 Reduction versus non-reduction

The term 'non-reductive physicalism' picks out a family of views which have been developed from the 1960s onwards. NRP is supposed to provide a 'third way' between two alternatives: reductive physicalism and emergentism. Standardly conceived, reductive physicalists hold that the psychological is nothing over and above the material in that mental properties are identical to physical properties. Emergentists hold that the psychological *is* over and above the material in that mental properties merely depend on physical properties in virtue of non-physical, fundamental 'laws of emergence'.[4] The alleged problem with reductive physicalism is that, despite its impeccable physicalist credentials, it fails to capture the distinctness of minds from matter. The alleged problem with emergentism is that, despite it capturing the distinctness of minds from matter, it has no physicalist credentials. The promise of NRP is supposed to be that it has the virtues of both of these views without the vices: a successful union of physicalism and the distinctness of minds from matter.

There have been a variety of NRPs offered over the last half century. Here, I detail the main versions of each, and explain how they differ, or are supposed to differ, from reductive physicalism. The chapter in outline:

I begin in Section 1.2 by outlining the view to which NRPists were primarily responding: reductive physicalism, which, according to the standard account, was originally defended by H. Feigl (1958), J. J. C. Smart (1959), David Lewis (1966), and David Armstrong

[3] (1989a).

[4] 'Emergent' (coined by Lewes 1874: 98) has various senses. J. S. Mill was the pioneer of the broader concept of emergence, if not the terminology (see Nagel 1961; McLaughlin 1992; Bedau & Humphreys 2008; O'Connor & Wong 2012). However, Mill's notion of diachronic causal emergence differs significantly from the concept of interest herein: synchronic non-causal emergence as most associated with C. D. Broad (see Mill 1843; Broad 1925). Other varieties of emergence are weaker: either strikingly similar to NRP, or the epistemological claim that emergent properties are those which any knower (either limited or omniscient), cannot predict or explain just given the emergence-base (see Bedau & Humphreys 2008; O'Connor & Wong 2012).

(1968), among others. Contemporary proponents include Jaegwon Kim (2005), Frank Jackson (2012), and Lawrence Shapiro and Thomas Polger (2012).

In Section 1.3 I detail ‘classic NRP’ which was initially motivated as a response to perceived problems with reductive physicalism, and was first proposed by Hillary Putnam (1960; 1967) and Jerry Fodor (1974), and defended by numerous others since.

I then discuss, in Section 1.4, Donald Davidson’s ‘anomalous monism’ (1980), which some have deemed a version of NRP, but which I reject as a version of NRP because it is not a physicalist view.

In Section 1.5 I examine ‘disjunctive physicalism’, advertised as a version of NRP by its proponents (Clapp 2001; Antony 1999, 2003) but which I argue is in fact a version of reductive physicalism as it accepts mental-physical property identity.

I conclude in Section 1.6.

1.2 Reductive physicalism

‘Reduction’ is a term of art which is used in a variety of senses in the philosophy of mind (see Shapiro & Polger 2012; van Riel 2014). As we will see, there is now something of a debate regarding the grounds on which we are to label a view ‘reductionist’ or not. In order to resolve such disputes it will be helpful to start again from the beginning, to examine the thought from which reductionism originally emerged.[5]

However, finding that source is not an entirely straightforward matter as there is some doubt as to when reductionism originally emerged. According to the standard story, the roots of reductive physicalism terminate in the work of U. T. Place (1956), Herbert Feigl (1958), and J. J. C. Smart (1959); although earlier theorists such as Rudolf Carnap (1932) and E. G. Boring (1933) also get honourable mentions (see Smart 2007: §1). However, given my characterisation of reductive physicalism, there would be not insignificant exegetical questions regarding the differences between Place on the one hand, and Feigl

[5] All uses of ‘reductionism’ and relevant cognates refer to mental-physical reductionism, unless stated otherwise.

and Smart on the other (ibid). To put such issues to one side, I treat Feigl and Smart as the starting point.

Smart is the philosopher with whom reductionism is most associated in the contemporary philosophical imagination. So, just as it would be wise to examine the work of, say, René Descartes, if we wished to ascertain the original intuition that motivated modern dualism, there is analogous wisdom in examining Smart's articulation of the original intuition that motivated reductionism.

So what is that intuition? Answer: for x to be reduced to y is for x to be *nothing over and above* y. Consider the following passage, where Smart introduces the much used phrase 'nothing over and above' into the philosophy of mind lexicon:

> the thesis [that sensations are brain processes] does not claim that sensation statements can be translated into statements about brain processes […] Nor does it claim that the logic of a sensation statement is the same as that of a brain-process statement. All it claims is that in so far as a sensation statement is a report of something, that something is in fact a brain process. Sensations are *nothing over and above* brain processes. Nations are *nothing 'over and above'* citizens, but this does not prevent the logic of nation statements being very different from the logic of citizen statements, nor does it insure the translatability of nation statements into citizen statements […] When I say that a sensation is a brain process […] I am using 'is' in the sense of strict identity (1959: 144-145; my emphasis).

Here, Smart is clearly stressing the importance of how things are *in the world* as opposed to how things are *in our theories*, *thought*, or *language*, and doing so with the language of identity: a sensation *is* a brain process–and in that sense the former is nothing over and above the latter.[6]

Somewhat confusingly, however, the 'nothing over and above' phrase now permeates the philosophy of mind literature, primarily as the name of the key intuition that *any* materialist position must do justice to, be it reductive or not. If one is to be a materialist about x, then one's view better make it look plausible to say that x is nothing over and

[6] Smart maintained this focus on ontology throughout his career, including his final thoughts on the debate (see Smart 2007).

above the material. Or so it is said. It is a curious feature of the history of the NRP vs. RP debate that the 'nothing over and above' phrase is something of a double-agent. Indeed, as will be discussed in this and subsequent chapters, the 'nothing over and above' phrase is not the only originally reductionist thing that NRPists have tried to claim as their own.

If we are to understand the original reductionist intuition as being an intuition about certain mental kinds of thing, like sensations, being identical to physical kinds of thing, like brain processes, then reductionism is best understood as a view which claims that mental properties are identical to physical properties.[7] And this, indeed, is how reductionism is standardly conceived, *pace* some recent attempts by NRPists to challenge this which we will consider below. Here, we will treat mental-physical property identity as at least sufficient for mental-physical reduction.

The first physical properties considered as candidates for identification with mental properties were neurological properties. Pain, it was thought, might be c-fibre activation. Granted, there were two predicates, one mental–'pain'–and one physical–'c-fibre activation'–but being in pain might just be nothing other than being in a c-fibre activation state. Different predicates, same property. Reductionists need not bet the farm on neurological properties, however. So long as mental properties can be identified with physical properties of some sort or other, reductive physicalism, in the sense employed herein, follows. This is the core tenet of RP:

(I) The reduction tenet: mental properties are identical to physical properties.

It is also worth emphasising the scope and variety of predicates reductionists have allowed for when formulating reductive statements. For example, Herbert Feigl, another of the original reductionists, discusses:

> the possibility […] of a one-many-ψ-ϕ correspondence [where 'ψ' denotes a mental predicate, 'ϕ' a physical predicate]. One could always formulate such a correspondence with the help of a general equivalence between statements containing single ψ-predicates on the one side and disjunctions of statements containing several and various ϕ-predicates on the other (1958: 391).

[7] Note that Smart uses 'reduction' and its cognates a number of times in 'Sensations and brain processes' (142, 148, 156).

Here's an example of how we might recast this thought in terms of properties. It might be that for every mental predicate there is a disjunctive physical predicate which is metaphysically co-extensive with it, where we understand the extension of a predicate as a class or a property. This opens up a route towards identifying mental and physical properties without the one-to-one matching between mental predicates and the familiar non-disjunctive physical predicates that the often presented textbook story of reduction would suggest. Many would find it implausible that pain = c-fibre activation, or any other single neurological or physical property for which we currently have non-disjunctive predicates. But it is important to note that reductionists themselves, from the very beginning of their project, foresaw the issue. Thus, finding claims like 'pain = c-fibre activation' implausible is not *ipso facto* to find reductionism implausible.

Despite the original focus on identity by reductionists, various historical contingencies resulted in Ernest Nagel's (1961) model of theoretic reduction becoming the standard account of reduction, particularly during the period when NRP rose to dominance in the 1960s and 1970s. One such contingency was that metaphysics was relatively unfashionable at the time. When metaphysics became more fashionable, property talk also became *de rigueur*. On Nagel's model, reduction is a relation between theories, where theory A reduces to theory B iff B, together with 'bridge laws', entails A.[8] Bridge laws connect the predicates of the reducing theory with predicates of the reduced theory. For example: 'x is F iff x is G' might express a bridge law where 'F' stands for 'pain' and 'G' for 'c-fibre activation'. The Nagelian model continues to play a significant part in contemporary discourse, despite prominent voices questioning its applicability (see Brooks 1994; Kim 1998, 2005; Melynk 2003, 2008; Shapiro & Polger 2012).

Here, we will be concerned with the ontological rather than the theoretic sense of 'reduction'. To justify this, it might be tempting simply to distinguish the theoretic and ontological senses–and stipulate that the latter is the only one of interest.[9] Some of my

[8] There is some debate about how Nagel's account of reduction ought to be interpreted (see Richardson 1979; Marras 2002; Klein 2009; Shapiro & Polger 2012). But that debate need not detain us here because all interpretations of Nagel agree that his account of reduction was an account of theory reduction, and it is that aspect of Nagel's view that is of primary interest.

[9] To be clear: by 'theoretic reduction' I mean the claim that, for the complete theory of the mental A and the complete theory of the physical B, all of the essential predicates in A can be derived, via bridge laws, from all of the essential predicates in B. I mean by 'ontological reduction' the claim that for every mental property F and physical property to G, if F=G then F is ontologically reduced to B. *Mutatis mutandis* for other theories and types of property.

interlocutors may well be satisfied with such a stipulation, but others may not. I have already given an argument for why the original sense of 'reduction' should be understood as ontological, but it will be worthwhile to consider some of the wider debate about which conception of reduction is to be preferred when framing the debate between RPists and NRPists. So bear with me while I explain why there are good reasons for preferring the ontological account to the theoretic one.

One might object to the ontological conception of reduction on the grounds that theoretic reduction is necessary for identifying mental and physical properties. So, if one wants to establish property identity, one needs to reduce the theories first, and so theory reduction is at least indispensable. I reject this argument for two reasons.

First, we have examples of property identifications arrived at in the absence of theory reduction, such as the discovery that water is H_2O.[10]

Second, the claim that to warrant identifying mental properties with physical properties it is necessary to be able to reduce psychological theory to physical theory seems plausible only if the theories in question are complete, and there are no epistemic barriers in place. Even if theories are complete, there remains the possibility that there may be epistemic limits on the subjects attempting the reduction, or other computational resource limits. In short, we might not be smart enough, or have fast enough computers. If we are not in the epistemically ideal scenario, then theoretic reduction is not required.

In response, one might point to purportedly successful theoretic reductions, like that of classical thermodynamics to statistical thermodynamics.[11] Without such a theoretic reduction, it might be argued, we would have no warrant for the ontological reduction of, say, thermal energy to kinetic energy.

[10] There is some debate in the philosophy of chemistry regarding how the phrase 'water is H_2O' is to be understood. The statement is perhaps best read as shorthand for 'pure water is a collection of appropriately interacting H_2O molecules' (see Weisberg, Needham, & Hendry 2011: § 4.5).

[11] Whether or not the reduction from classical thermodynamics to statistical thermodynamics is successful is controversial (see Sklar 2009: § 6). Indeed, whether or not any reductions in Nagel's sense are possible is controversial (see Kim 1998, Shapiro & Polger 2012: 273-275). At the present moment, I am not arguing in favour of either side of these debates. Instead, I am pointing out why one need not be a Nagelian to be a reductionist.

I grant that as a matter of fact ontological reductions in scientific history have sometimes been guided by theoretic reductions. I think that there are at least two significant ways in which this has occurred.

First, theoretic reduction has been used to settle the issue of whether or not there should be ontological reduction at all. But even in such cases this does not entail that theoretic reduction is essential for ontological reduction, or that theoretic reduction is the fundamental concern when we are in pursuit of reductive goals. An analogy: astronomy might demonstrate that the Morning Star is the Evening Star, but one could accept that there is a single planet without astronomy (via superhuman eyesight, say), and what really matters is that there is a single planet, not the theoretic derivation of that fact from star charts and a model of the solar system (say).

Second, theoretic reduction has been used to settle the issue of exactly which entities are identical to which other entities. An analogy: one could be content with the hypothesis that the Morning Star and the Evening Star are the same planet, but require astronomy to determine precisely which planet that is. Similarly, in the mind-body case one could be content with the hypothesis that mental properties are physical properties but leave it open to the relevant sciences to figure out just which properties are identical to which.

Thus, with regards to the argument that theoretic reduction is necessary for ontological reduction, the bottom line is that given the narrow range of cases in which it is *a priori* plausible to assert that theoretic reduction is essential for property identification, and the *a posteriori* cases of property identification in the absence of theory reduction, we should reject the claim that mental and physical properties can be identified only if the relevant theories can be reduced.

Moreover, there is a powerful counter-argument against theoretic reduction, namely that ontological reduction is necessary for it (although whether one should ultimately accept the argument is a complex question, as we will see). The argument is that the Nagelian model won't work unless the bridge laws are understood as identity statements, and thus that property identity is in fact necessary for theoretic reduction–indeed, this is a commonly accepted claim (see Batterman 2012: §1). This is often motivated by considering how actual reductions in science seem to have proceeded. Consider classical optics and the theory of electromagnetic radiation. The reduction of the former to (part of) the latter did not proceed merely by finding *correlations* between true applications of the relevant theoretical terms, but rather by *identifying* light wave properties with some electromagnetic properties (Sklar 1967: 120).

Nevertheless, one could always hold that the identification in such cases was a step too far: perhaps the correlation (or some other relation short of identity) is all we are really justified in believing. What is often appealed to in order to motivate the claim that identification is *required* is a perceived need to explain the holding of bridge laws. The argument proceeds as follows. If it is empirically discovered that x is F iff x is G, then the task remains to explain why one property is instantiated iff the other is. Given that these are *bridge* laws, one cannot appeal to the explanatory resources of either the reduced or the reducing theory as bridge laws by their very nature cannot be part of either. Also, appeal to a further law that is not part of either the reduced or reducing theory to explain the bridge law invites regress. If the holding of bridge law L_1 is explained by appeal to the holding of a further bridge law, L_2, then we face the task of explaining the holding of $L_{2;}$ and if we explain the holding of L_2 by appeal to a further bridge law, L_3, then we face the task of explaining the holding of L_3; and so on. The supposedly simple solution, therefore, is to identify F and G, because, the claim goes, this explains the holding of the bridge law and provides an explanatory terminus.

If this argument works, then even if one is a determined adherent of theoretic reduction, one would have to accept that the model only works when interpreted as mapping property identities. I think that the argument does count strongly against theoretic reduction, but with two major caveats, which I will only touch on here, as I discuss this further in subsequent chapters where the issues have particular importance for the dialectic.

The first caveat is that the holding of identity relations *can* require justification, if there is significant room for doubt regarding the truth of the identity statement reporting the holding of the identity relation. And in the mind-body case, infamously, we have doubt aplenty. Thus, the extent to which an identity claim is an ideal explanatory terminus can be questioned.

Consider how Ned Block and Robert Stalnaker try to rule out that identity statements can require justification:

> If we believe that [temperature] is correlated with but not identical to molecular kinetic energy, we should regard as legitimate the question of why the correlation exists and what its mechanism is. But *once we realize* that [temperature] *is* molecular kinetic energy, questions like this will be seen as wrongheaded (1999: 24; emphasis added).

The problem with Block and Stalnaker's claim can be seen in the phrase I have italicised: 'once we realize'. To see this, distinguish the following two points: (i) whether it makes sense to ask why A=B; (ii) whether it makes sense to ask why we should believe that A=B. Block and Stalnaker are correct if we read them in the sense of (i)–everything is identical to itself (let us hope), and asking why something is identical to itself is certainly wrongheaded. But Block and Stalnaker are incorrect if we read them in the sense of (ii)–all sorts of identity claims can be justified, or not, in this sense. Indeed, in the mind-body debate we face a particularly clear example of a case in which identity claims require justification. And the 'once we realise' phrase rather obscures the distinction between (i) and (ii), as it simply assumes that (ii) has been dealt with, hence all that remains is (i)–which is trivial, as they point out.[12]

Thus, one could accept that if mental-physical identity holds, then the bridge law regress problem is solved, while denying that simply appealing to identity blocks questions about why one should accept that the identity in question holds. This is an important point that we will return to in subsequent chapters.

The second caveat is that it is not immediately obvious that mental-physical identity is necessary for solving the bridge law regress problem. This is because there is another option for explaining the holding of bridge laws without appealing to identity: that the bridge laws could be entailed by the nature of the relevant relata (in something like the way that *is taller than* is entailed by Jill's being six feet tall and Jack's being five feet eight inches tall). If that were the case, then questions about why the bridge laws hold could seem nearly as wrongheaded as questions about why A=A. This brings us face-to-face with a major background issue in all of this: namely whether the bridge laws in question are *a*

[12] As it happens, Block and Stalnaker state that 'identities don't *have* explanations (although of course there are explanations of how the two terms can denote the same thing)' (1999: 24). But this seems to understate the issue: it's not that we want to know why we should accept that 'pain' and 'c-fibre activiation' denote the same property, *per se*, it's that we want to know why we should accept the claim that pain is identical with c-fibre activation, when it appears that pain isn't identical with c-fibre activation or any other physical property with which we are familiar.

posteriori or *a priori*. When this debate began, the assumption was that the bridge laws must be empirical, and with this assumption in mind the force of the regress problem seems clear: once you empirically discover the first bridge law, and accept the demand to empirically explain its holding, you can only do that by discovering another bridge law, and the regress has begun. But if, given the nature of the relata, the bridge law is entailed, then the regress can't get started. I postpone further discussion of these issues until Chapters 2 and 3 (Section 2.3 and 3.2, respectively), where I consider this point again in relation to the explanatory demands faced by NRP as a physicalist position. For present purposes, simply note that any proponent of the regress problem needs to address the option we have noted here in the second caveat.

We turn now to consider one final way of defending theoretic over ontological reduction that involves pointing to apparent differences between reduction and identity. Consider some of the logical properties of the *reduces* relation. Plausibly, if A reduces to B, then A stands in an asymmetric and irreflexive relation to B: A reduces to B, but B does not reduce to A, or to itself. We reduce water to H_2O, we do not reduce H_2O to water, nor water to itself. On the other hand, A *ontologically* reduces to B if A is identical to B, but identity is symmetric and reflexive. If water is identical to H_2O, then H_2O is identical to water, and everything is identical to itself. But if that's the case, then the conception of ontological reduction is incoherent: no relation can be both asymmetric and symmetric, never mind reflexive and irreflexive.

The way out of this quandary, the defence of theoretic reduction continues, is to abandon the ontological conception of reduction, in favour of theoretic reduction, which does not rely on identity to make good on our talk of reduction, and so is not committed to reduction being reflexive and symmetric.

There are interesting and complex solutions on behalf of ontological reduction that I will not delve into the detail of here (for an extensive discussion, see van Riel 2014). I think that the problem can be dealt with relatively straightforwardly. First, note that one might try to argue that theoretic reduction faces a similar problem: if bridge laws are of the form 'A iff B', then theoretic reduction is symmetric too, thus, the thought would go, the defence of theoretic reduction fails because the problem it relies on is not just a problem for ontological reduction, *per se*; it's a more general problem with giving an account of reduction which fully accommodates our somewhat conflicting intuitions about what reduction is.

However, this misses an important point: even if bridge laws are biconditionals, there may be some important derivations that only go in one direction, and thus account for the apparent asymmetry. Consider: perhaps one of the most plausible cases of reduction involves temperature. Grant that there is a bridge law of the form 'A iff B' where 'A' stands for a temperature law of classical thermodynamics, and 'B' stands for what we discover is lawfully correlated with that law in statistical thermodynamics: which for a gas will involve mention of mean molecular energy; for a solid mention of mean maximal molecular energy; and for plasma mention of kinetic energy per particle. Thus, given the biconditional, it might seem as if there is symmetry, and there is. However, there is also a significant unidirectional derivability here: just given the temperature law of classical thermodynamics, even if one is in possession of the bridge law, one can only derive the entire disjunctive law which mentions gases, solids, and plasmas. However, just given, say, a relevant law about mean molecular energy in gases, plus the bridge law, one can derive the temperature law of classical thermodynamics. Thus, there can be symmetry with directionality, and the intuitive asymmetry of reduction can be accounted for consistently with bridge law symmetry.

Importantly, we can use the same pattern of reasoning when defending the ontological conception of reduction. Namely, note that directionality of the sort described above can accommodate the intuitive asymmetry of reduction consistently with the purportedly problematic features of identity: temperature might be identical with the disjunction: mean molecular energy in a gas, or mean maximal energy in a solid, or kinetic energy per particle in plasma (I address concerns with using 'disjunction' in this way below in Section 1.5). But this is consistent with the unidirectional derivability we noted above, and thus the intuitive asymmetry of reduction can be accounted for consistently with identity.

And note how naturally this account fits with the concern with which reductionism began: how things are in the world, rather than how things are in our theories, thought, or language. Although also note that what I do not intend to do here is rule out that there are other senses of 'reduction' which may be interesting and useful in other contexts. My contentions relate only to the sense of 'reduction' that should matter to us when trying to solve the mind-body problem.

To sum up: the ontological conception of reduction does justice to the original reductionist intuition, and can resist the objections of those who would prefer the theoretic conception of reduction. Those objections were that theoretic reduction is necessary for ontological reduction and that talk of 'ontological reduction' is incoherent. We saw how those objections failed and how theoretic reduction faced a problem of its own: explaining bridge laws without regress. This lent further support to the claim that ontological reduction should be our primary concern. Of course, our theories ought to correspond to the facts, and it is natural to expect any reductive relationship between entities in the world to be reflected in our ideal theories. But this is a secondary concern. Reduction is fundamentally a matter of the reduced being nothing over and above the reducer; a matter, in other words, of the reduced being identical to the reducer.

1.3 Classic non-reductive physicalism

Classic non-reductive physicalism achieved something relatively rare in philosophy: being held by nearly everyone who matters. Throughout much of the 1970s, 1980s and 1990s an 'anti-reductionist consensus' was achieved among physicalists (Block 1997: 107), before reductionism began something of a revival.[13] NRP remains the dominant view, however, at least in so far as there are more people in the philosophy of mind who self-ascribe 'non-reductionist' than there are who self-ascribe 'reductionist'. As we will see, though, there is a question to be asked about how accurate some of these self-ascriptions are.

NRP grew out of objections to reductionism first put by Hilary Putnam via an analogy between computational properties and mental properties. Indeed, if any one event marked the beginning of the schism between reductive and non-reductive physicalists, then it was the third annual New York University Institute of Philosophy conference, which occurred in May 1959 and was attended by many leading figures of the period. In the same year Smart published the most famous single statement of mental-physical reductionism: 'Sensations and brain processes'. The proceedings of the conference were published the following year, and included a paper by Putnam called 'Minds and machines'–which is a

[13] As indeed, did anti-physicalist views – particularly following Chalmers (1996).

foundation document of NRP of a stature equal to Smart's paper as a foundation document of RP.[14]

The basic thought that Putnam articulated in 'Minds and machines' is that computational properties can be *realised* by different physical properties, such as mechanical or electronic properties, and that mental properties could be computational too, and thus also be realised by different physical properties. For example, given that one can describe the computational property of *calculating division* in functional terms, i.e. in terms of causes and effects, then different physical properties might 'play the functional role' of calculating division by producing the relevant effects in response to the relevant causes. Similarly, if thought is functional, i.e. if what it is to think is simply to be in a state that plays the right kind of role, then different physical properties could play the various roles of thought. Hence NRP's defining mantra: 'the mental is *multiply realised* by the physical.'

Interestingly, Putnam did not employ his original version of this argument directly against RP, but rather as part of an extended critique of the mind-body problem as a genuine problem.[15] Nevertheless, once such quasi-positivist thinking (for that is what it was) fell away, Putnam, and indeed most of his contemporaries, came to believe that the falsity of reductive physicalism follows from the argument. As Putnam stated in 'Minds and machines':

> the 'logical description' of a Turing machine does not include any specification of the physical nature of [the machine's internal] 'states' or indeed, of the *physical nature* of the whole machine. (Shall it consist of electronic relays, of cardboard, of human clerks sitting at desks, or what?) In other words, a given 'Turing machine' is an *abstract* machine which may be physically realized in an almost infinite number of different ways […] It is interesting to note that just as there are two possible descriptions of the behavior of a Turing machine–the engineer's structural blueprint and the logician's 'machine table'–so there are two possible descriptions of human

[14] As C. V.Borst states, about Smart's paper, 'Since its publication in 1959 this relatively brief article has been veritably echoing through the pages of the journals in the form of discussion which it has engendered, and must, in addition, surely be near to creating a record as the most frequently reprinted article' (1970: 16).

[15] As Putnam states: 'The moral, I believe, is quite clear: it is no longer possible to believe that the mind-body problem is a genuine theoretical problem, or that a "solution" to it would shed the slightest light on the world in which we live. For it is quite clear that no grown man in his right mind would take the problem of the "identity" or "non-identity" of logical and structural states in a machine at all seriously not because the answer is obvious, but because it is obviously of no importance what the answer is. But if the so-called "mind-body problem" is nothing but a different realization of the same set of logical and linguistic issues, then it must be just as empty and just as verbal' (1960: 175).

> psychology. The 'behavioristic' approach […] aims at eventually providing a complete physicalistic description of human behavior, in terms which link up with chemistry and physics. This corresponds to the […] physicist's description of a physically realized Turing machine. But it would also be possible to seek a more abstract description of human mental processes, in terms of 'mental states' (physical realization, if any, unspecified) (1960: 159-161).

Once this initial functionalist thought was had, the direct attack on reductionism followed and intensified into a critique that RP was obviously, intuitively chauvinistic:

> If we can find even one psychological predicate which can clearly be applied to both a mammal and an octopus (say 'hungry'), but whose physical-chemical 'correlate' is different in the two cases, the brain-state theory has collapsed. It seems to me overwhelmingly probable that we can do this. Granted, in such a case the brain-state theorist can save himself by *ad hoc* assumptions (e.g. defining the disjunction of two states to be a single 'physical-chemical state'), but this does not have to be taken seriously (Putnam 1967: 165).

By 'brain-state theory' Putnam is referring to the textbook version of reductionism whereby pain = c-fibre activation (or some other neurological property). As we have seen, to identify brain-state theory with physical reductionism would be a mistake because physical properties other than just neural properties could be identical with mental properties. It is an interesting feature of the history of NRP that it was in large part motivated as a response to the most implausible version of the view it opposed.

The other main position that NRP was motivated in response to was the positivist Unity of Science movement, the *locus classicus* of which is Oppenheim and Putnam's 'Unity of Science as a Working Hypothesis' (1958).[16] The unity of science, it was argued, consisted in the relationship between the different ordered 'levels' of reality being reflected in the structure of scientific disciplines. Each science had its own 'level', with the 'bottom' and most general level being occupied by microphysics, the next level 'up' by chemistry, then biology, psychology, and sociology (to simplify greatly). This model was supposed to reflect the mereological relationships between the entities that were the concern of each

[16] This is the same Putnam who we identified above as the founding father of NRP, thus Putnam ended up arguing, later in his career, against the unity of science position he defended with Oppenheim in 1958.

science. The elementary particles (microphysical level), were parts of molecules (chemical level), which were parts of cells (biological level), which were parts of individuals (psychological level), which were parts of societies (sociological level). Oppenheim and Putnam held that each level micro-reduced to the level below, in that the mereological relationships between the entities at adjacent levels were sufficient to licence reduction of the laws of the reduced level to the laws of the reducing level.

Whatever the merits of this picture, the standard reading of it by NRPists treated the reduction claim as requiring Nagelian bridge laws, and it was largely on this basis that the *dis*unity of science was argued for.[17] That is, when NRPists criticised the Unity of Science movement, they did so by arguing against the possibility of bridge laws holding between any of the sciences in question. Fodor provides the definitive example:

> Gresham's law[[18]] says something about what will happen in monetary exchanges under certain conditions. I am willing to believe that physics is general in the sense that it implies that […] any event which falls under Gresham's Law […] has a true description in the vocabulary of physics and in virtue of which it falls under the laws of physics. But banal considerations suggest that a description which covers all such events must be wildly disjunctive. Some monetary exchanges involve strings of wampum[[19]]. Some involve dollar bills. And some involve signing one's name to a cheque. What are the chances that a disjunction of physical predicates which covers all these events [...] expresses a physical natural kind? (1974: 103).

We will return to the question of how plausible the disjunction option is below. But for present purposes it is important simply to note that NRP grew out of a rejection of RP, the latter of which was understood as requiring bridge laws which, according to NRPists, obviously didn't hold. This was originally combined with a functionalist view of mental states such that they are analogous to computational states, which also offered an

[17] A large part of the reason for the assumption of bridge laws was that the model of explanation that was assumed was the deductive nomological (DN) model of Carl Hempel and Paul Oppenheim (1948), which maintains that phenomena are explained by deducing them from natural laws. According to the (DN) model, to give an explanation of a phenomenon is to subsume it under natural law. That subsumption involves the phenomenon being deduced from the laws plus boundary conditions; laws themselves are derived from other more fundamental laws.

[18] Named after the 16th century Royal adviser, Sir Thomas Gresham. The maxim states that 'bad money drives out good (but not vice versa). The law was first observed to hold because of the common practice of counterfeiting, or simply 'clipping' gold or silver coins to remove small amounts of precious metal. The bad money (counterfeit or clipped coins) was quickly passed on, whereas the good money was not and thus was driven out of circulation (Law & Smullen 2008).

[19] Small beads typically used by Native Americans.

explanation of why the bridge laws couldn’t hold: mental states are multiply realised. It was out of these ideas that NRP arose.

With this in mind, we can now outline the classic NRP metaphysic that solidified around these ideas.

In the broadest terms, there are three core tenets of NRP:

(1) The nothing above tenet: Minds are nothing over and above matter.
(2) The distinctness tenet: Minds are distinct from matter.
(3) The causation tenet: Minds cause matter to move.

The many proponents of NRP differ with each other over just how these tenets are to be interpreted. The critics of NRP differ with the proponents usually by denying that, given some interpretation or other, (1) is consistent with (2), or that (2) is consistent with (3).

The standard minimal interpretation of these tenets is as follows (these are necessary conditions):

(1*) The nothing above tenet: Mental properties supervene on physical properties.
(2*) The distinctness tenet: Mental properties ≠ physical properties.
(3*) The causation tenet: Mental properties are causally relevant for physical properties.

Let’s consider each of these in turn.

First, what is supervenience? At heart, a supervenience claim is very modest and it is difficult to make much progress in the philosophy of mind without adding to it in some way. Indeed, as we will see, the main challenge facing NRP is the question of just what needs to be added to the supervenience claim in order for NRP to be viable. Minimally, to say that ‘A supervenes on B’ is to say that ‘there cannot be an A change without a B change’. This states a pattern of co-variation between A and B, which are taken to stand in a transitive, reflexive, non-symmetric, synchronous, non-causal relation with some unspecified degree of modal force. This is compatible with B not entailing, or explaining, or reducing to, or ontologically depending on A; and vice versa. But even though few in the philosophy of mind use ‘supervenience’ in this minimal sense, one encounters in the literature technical definitions of supervenience which do not explicitly move much beyond the minimal conception, immediately followed by uses of ‘supervenience’ in a less modest sense (e.g. Kim 1998: 9-11). That less modest sense is: ‘ontologically depends on’ intended to refer to a transitive, irreflexive, asymmetric, synchronous, non-causal relation with modal force, of which the following is true (where ‘A’ and ‘B’ stand for the domain of mental and physical properties respectively):

A depends on B only if: *necessarily*, if anything x has some property F in A, then there is at least one property G in B such that x has G, and, *necessarily*, everything that has G has F.

Nearly everyone in the philosophy of mind who uses ‘supervenience’ does so to refer to an ontological dependence relation–hence the talk of ‘*sub*vening’ properties, and ‘supervenience *base*’. Such talk is supposed to reflect the thought that the mental is somehow ‘fixed’ by, and relies for its existence on, the physical–but not vice versa. However, none of this is implied by ‘supervenience’ in the original sense. I will not follow the somewhat loose practice of using ‘supervenience’ to mean ‘ontological dependence’. Instead I will use ‘dependence’.

The question is then to specify the modal force of dependence, the two main options being either nomic or metaphysical. Although in the above definition of dependence many NRPists would read the first necessity operator as holding with nomic force, the second with metaphysical.

Closely related to the claim that the mental depends on the physical is the claim with which NRP began: the mental is realised by the physical. A large amount of the effort that has gone into the development of NRP has gone into developing an account of what it is for a mental property to be realised by different physical properties. Indeed, we will consider a family of views which attempt to do just that when we assess immanentism in Chapter 3. It should be noted that talk of ‘realisation’ is sometimes treated as equivalent to talk of ‘dependence’. But NRPists often mean something much meatier when they use ‘realisation’, and classic NRPists developed an account of realisation that became the standard version.

According to that account, which grew out of the functionalist model that Putnam drew the very first sketch of in ‘Minds and machines’, mental properties are defined in terms of their causal role, where this is understood as the mental property being identical to the property of being some property or other that plays the relevant causal role (see Putnam 1970). This definition operates as the name for a higher-order relational property. The world can then be empirically investigated to see which physical properties actually perform the causal role of the mental property. To use the standard toy example, the property of pain might be defined as follows: being in pain $=_{\text{def.}}$ having some property or other which is caused by bodily injury and causes wincing or moaning. We can then look inside humans to see what physical property causes wincing and moaning in response to bodily damage, and find that being in a c-fibre activation state is that property. In that case, according to this account, we would say that c-fibre activation realises pain in humans. Crucially, though, when looking inside other species, or within the same species at different times, we find (or could find), the claim goes, different physical realisers of pain. To use Lewis’ evocative example, perhaps pain in Martians is realised by the inflation of tiny cavities in their feet.

Second, the non-identity condition on mental and physical properties is of course necessary for the denial of RP, but some NRPists go further and also deny that mental events are identical to physical events, and in that way are dualists about both properties and events. This is probably the preferred position among most NRPists today, but, classically, most NRPists held event monism: that mental events are physical events–and property dualism–mental properties are not physical properties. This latter view was often called the ‘token identity theory’–in contrast to the type (i.e. property) identity theory of Smart and the other RPists. The token identity theory was the dominant version of classic NRP.

Third, it is seen as the most basic condition on the claim that minds cause matter to move that mental properties are causally relevant for physical properties. There has been much ink spilt over what it is for one property to be ‘causally relevant’ for another. But here’s the basic idea: a common *façon de parler* in the literature is to talk of an effect being caused ‘in virtue of’ some properties of the cause, but not others. For example: the ball’s crossing the net was caused in virtue of the speed and angle of the racket swing, not that the racket swing happened on a Tuesday. We will see an example how one might spell out the ‘in virtue of’ talk when we discuss Davidson, below. But for present purposes, this basic idea will suffice.

One potential complication and point of dispute among NRPists is whether, in order to do justice to the claim that minds cause matter to move, one not only has to show that mental properties are causally relevant for physical properties, but also whether one has to show that mental events cause physical events. If one is an event monist, then the answer is an easy one: mental events just are physical events. But if one is an event dualist, then the issue becomes more difficult, simply by virtue of the doubling of the explanatory task.

A quick note here about the terminology: the distinction between event and property causation is often put in terms of 'causal efficaciousness' versus 'causal relevance'; events are efficacious, properties relevant. For example, the event of the earthquake was causally efficacious with respect to the event of the building toppling over, but it was the force (property) of the earthquake and the weakness of the building (property) which were causally relevant for the speed (property) with which the building fell. One also sometimes encounters a distinction between causal relevance understood as a kind of (at least partly) mind-dependent causal explanation versus causal efficaciousness understood as a kind of mind-independent causal relation.

Indeed, we turn now to consider a potential major variant of NRP distinct from classic NRP, which faced a major problem framed in terms of the distinction between event efficaciousness and property relevance: anomalous monism.

1.4 Anomalous monism

Donald Davidson's position is worth considering as a standalone theory because he provides a good example of the kind of theorist who is arguably an NRPist in the broadest sense, and has certainly been claimed as an NRPist by some (e.g. Antony 2007: 146; Baker 2009: 110-113), but yet whose characterisation as an NRPist is controversial. I think that Davidson is not an NRPist because he is not a physicalist, and if we re-interpret his view to make him a physicalist, his view collapses into classic NRP.

Davidson's view differs from classic NRP in two key respects. First, his view of the causal status of mental properties. Second, his view of supervenience. Let us consider each in turn.

Begin by noting what puts the monism in ‘anomalous monism’–the claim that all causally interacting mental events are identical to physical events. This can be taken to mean that every event with a mental property is identical to an event with a physical property. This interpretation of Davidson, although relatively common in some parts of the literature (e.g. Kim 1998; 2012), is false (see Gibb 2006). Davidson’s ontology did not include properties in anything like the sense used here. This makes it difficult to engage with Davidson’s view using the framework standardly assumed in the philosophy of mind, so here I will follow the practice of others who, when engaging with the Davidsonian position, do so by taking his talk of properties (of which there was quite a bit) at face value. We will return to this point in Chapter 4 (Section 4.1) where we draw a parallel between Davidson’s denial of the existence of properties, and the possibility that some NRPists might deny that there are property instances. But for present purposes, let us simply note that it is possible that Davidson be read as committed to the existence of properties, this reading is relevant to my survey of NRP, thus justifying his inclusion here. So, here, we shall follow those who grant that Davidson’s many references to ‘properties’ should (or must) be taken literally.

Given this interpretation, Davidson’s ontology is monist in so far as events are basic and can possess either mental or physical properties. To that extent at least, Davidson’s view can look a lot like a kind of token identity theory. One of the most interesting aspects of Davidson’s view is that his monism is derived from further principles that had previously been considered to be incompatible, and certainly incompatible with monism. Those principles were:

Interaction Principle: mental events interact causally with other mental and physical events.

Cause-Law Principle: events which are causal relata must fall under exceptionless laws.

Anomalism Principle: mental events as such do not fall under exceptionless laws.

Non-Anomalism Principle: only physical events as such do fall under exceptionless laws.

The surprising conclusion that Davidson drew from these principles was that if x is a mental event, then x must be a physical event: *monism*.

That Davidson's identity claim only holds between tokens and not types is necessitated by his denial of nomic connections between the mental and the physical. If the truth of the anomalism and non-anomalism principles is granted, then the lack of strict laws, and therefore nomic connections between the mental and the physical, follows.[20]

But given the interaction principle, and the cause-law principle, causally interacting mental events must have some property amenable to the inclusion of those events in strict laws; and, Davidson maintains, fundamental physical properties are the only properties included in such laws, therefore every event with a mental property must be identical to an event with a fundamental physical property; hence event token identity.

One of the key ideas behind Davidson's entire metaphysic of mind is that causal relations between events can be analysed as instances of lawful regularities. Consider the following, simplified version of the view:

C is the cause of *E* if there is a law which subsumes them, $C \neq E$, and *C* is spatiotemporally proximate to *E*.

A law L subsumes events *C* and *E* iff there are predicates F and G such that *C* satisfies F and *E* satisfies G, and L entails:$\forall x(Fx \rightarrow Gx)$; and '$\forall x(Fx \rightarrow Gx)$' is not a logically necessary truth. For example, the law: all pure water boils at 100 degrees centigrade in Earth's atmosphere subsumes the events *this pure water was heated to 100 degrees centigrade at t* and *this pure water boiled at t*. Thus, nomic subsumption is a species of nomic implication between sentences containing correct event predications.

This Davidsonian metaphysic gives us the following model of mental causation. Consider two events, *C* and *E*, such that *C* causes *E*, and F, G, H are the properties of *C*, and I, J, K, are the properties of *E*. Further, F and I are mental properties of *C* and *E* which are referred to by the predicates in a true folk psychological generalisation about *C* and *E*, and H and K are microphysical properties which are referred to by the predicates in a strict physical law which states a causal regularity. Thus, according to this metaphysic, *C* causes *E* in virtue of

[20] This is because the anomalism principle not only rules out the existence of exceptionless laws of succession in which mental predicates figure in the consequent, e.g. '$P_1 \rightarrow M_1$', or '$M_1 \rightarrow M_2$'; but because Davidson also denies the possibility of any exceptionless laws which include mental predicates, including the kind of psychophysical bridge laws of the form '$P_1 \leftrightarrow M_1$'.

the strict physical law which refers to H and K, not in virtue of the generalisation which refers to F and I. Consequently, only the physical properties are causally relevant.

For example, if F is being an experience of a tree, and I is being a belief that there is a tree, then, according to this understanding of the Davidsonian view, the experience would not cause the belief in virtue of the mental properties of those events; in virtue, that is, of the former event being an experience of a tree, and the latter event being a belief that there is a tree. Instead, *C* would cause *E* in virtue of H and K. What is important to note is that although Davidson has given an account which accommodates, to some extent, the causal efficacy of mental events, those events are not efficacious in virtue of their mental properties; in virtue, in other words, of their mentality. In this way, Davidson's view appears to entail the causal irrelevance of mental properties. Thus, Davidson does not adhere to the third tenet of NRP.

Davidson responded to this charge by insisting that one cannot properly speak of one event causing another 'in virtue of F rather than G', because, on his view, causation is simply a relation between events–properties don't come into it (Davidson 1993: 6). This response was partly justified by Davidson denying that are properties in the relevant sense. This response was viewed as pushing the nuclear button: throwing out an entire ontological category, properties, to save a particular position in the philosophy of mind. Indeed, Davidson's refusal to try to do justice to the intuition that events cause what they do in virtue of their properties was a significant waypoint along the decline in interest in anomalous monism that occurred during the last 25 years. This point will prove to be pertinent when we consider whether NRP can really claim to be a physicalist view in Chapter 4.

Whether or not Davidson fails to adhere to the third tenet of NRP, there are good grounds for thinking that he fails to adhere to the first tenet. This brings us to our second, and final point. Davidson wasn't an NRPist because Davidson, who introduced the notion of supervenience into contemporary philosophy of mind (1970: 75), opted for weak supervenience (1993: 4n4). But this is insufficient for physicalism for reasons I will now explain.

Recall that, in the original sense of 'supervenience', the relation is reflexive and non-symmetric, although when employed in the philosophy of mind a different sense of ‘supervenience’ is intended–in particular an irreflexive, asymmetric sense. After Davidson introduced supervenience into the philosophy of mind, a whole industry grew up around the project of getting to grips with the relation and its various senses–particularly those which seemed relevant in the philosophy of mind. Out of that project a number of important distinctions emerged, one of which is relevant here: weak versus strong supervenience.

Weak supervenience only holds intra-world: A weakly supervenes on B iff there is no world such that any individuals in that world are B-*in*discernible but A-*dis*cernible. In contrast, A strongly supervenes on B iff, given some range of worlds, there are no individuals that are B-*in*discernible but A-*dis*cernible whether they are in the same world or not.

For various reasons that we need not detain ourselves with here, Davidson has committed himself to the claim that the mental weakly supervenes on the physical. But this leads to three major problems.

First, adopting weak supervenience invites a difficult explanatory question: why would mental properties weakly supervene without strongly supervening? Why, in other words, would it be the case that supervenience holds intra-world, but not inter-world? It is far from clear what a plausible answer would look like, particularly given that the answer cannot entail that strong supervenience holds (see Blackburn 1984). Compare: if someone holds strong supervenience, and they are asked why two individuals in a particular world within the relevant range are mentally indiscernible if they are physically indiscernible, that someone can explain this by appeal to the strong supervenience claim–it’s necessarily so, they can say. They face further questions about why it is necessarily so, but they have at least discharged the initial explanatory burden. But weak supervenience theorists can’t even take this first explanatory step.

Second, it is not clear how consistent Davidson’s weak supervenience claim is with his denial of nomic connections between the mental and the physical, because weak supervenience, it would seem, just is an intra-world exceptionless law.

Third, and most importantly of all, weak supervenience is insufficient for physicalism, because, if one is a physicalist then one must hold that the physical facts fix all the facts.

But, if one holds that the mental only weakly supervenes on the physical, then there are worlds which are physically indiscernible from the actual world, but mentally discernible– and hence the physical facts do not fix all the facts. In other words, the first tenet of NRP, the mental is nothing over and above the physical, requires something stronger than weak supervenience.

Thus, anomalous monism is not version of NRP. We might re-interpret anomalous monism, of course, as some have, but the most obvious ways of doing so, such as strengthening the supervenience claim, would only bend anomalous monism into a version of classic NRP.

One further NRP variant of interest remains: disjunctive physicalism.

1.5 Disjunctive physicalism

Around the turn of the century, some NRPists began defending a position that had, up until that point, been viewed as a version of reductive physicalism (see Antony 1999, 2003; Clapp 2001). In short, they argued that mental properties were to be identified with a disjunction of physical properties, yet this picture was not reductionist because the ‘autonomy’ of the mental could still be preserved given this metaphysic. We will now see why this is reductionism in (rather poor) disguise.

The first thing to get clear on is that there cannot be disjunctive properties any more than there can be disjunctive grand pianos, so let us begin by clarifying what ‘disjunctive property’ means. A predicate is properly disjunctive iff there is more than one disjunct, each disjunct designates a real property, and each disjunct designates a distinct property. Then, F is a *disjunctive property* iff F can be designated by a properly disjunctive predicate (here I follow Clapp 2001: 123).

In what follows, I will talk of ‘Fs’ and ‘Gs’ being ‘co-extensive’. To be clear, I take the extension of a singular term to be its referent: the extension of ‘Alex Salmond’ is Alex Salmond. The extension of a predicate is a class or a property: the extension of ‘is Scottish’ is the class of Scottish things, or the property of Scottishness (I leave this open here because I wish to argue below that disjunctive predicates can pick out properties, but do not want to beg the question in my definition of 'extension').

Now for *disjunctive physicalism*: for any mental predicate there will be a (perhaps infinite) disjunctive physical predicate which is at least nomically co-extensive with it. If this is true, then at the very least it blocks the classic NRP arguments against mental-physical property identity: the multiple realisation argument, and the related argument against bridge laws. This opens the way to identifying mental and physical properties – on parsimony grounds, say. Indeed, NRP defenders of disjunctive physicalism cede that mental properties are identical to physical properties, although for different reasons (Antony 2003: 9; Clapp 2001: 114).[21] Clapp reasons that nomic extensiveness entails theory reduction, the latter of which is sufficient for property identification; as he states:

> The classic reductionists, or 'type-type identity theorists,' such as U.T. Place and JJ.C. Smart, make it abundantly clear that a primary motivation for their theory is the ontological parsimony that results from identifying mental states with physical states. In the sense in which reduction is relevant to the mind-body problem and the tenability of NRP, the reduction of one theory T_1 to another theory T_2 collapses the ontological commitments of T_1 to those of T_2; if T_1 is reduced to T_2, then one who endorses both theories is committed to the existence of only the entities posited by T_2 (2001: 114).

But, despite the concession of mental-physical property identity, disjunctive physicalism is non-reductive because, it is claimed, given the massive disjunction that will be required to produce a physical predicate co-extensive with any mental predicate, it will still be the case that, unless one is God-like, mental predicates will not be deducible from or replaceable with physical predicates–and this is all that any defender of NRP requires.

Or so the argument goes. Given my characterisation of reductive physicalism (M reduces to P if M=P), disjunctive physicalism is a version of it. And it is not difficult to trace a tradition of thought back to the original non-reductionists, and reductionists, according to

[21] Antony is somewhat more circumspect, stating: 'I've been intimating that there's really no reason for defenders of MR to resist the identification of mental properties with disjunctive physical properties, so let me come completely clean. What I really think is that there is no such thing as a ''disjunctive property''—rather, there are only disjunctive predicates. Moreover, I think that properties are, in themselves, neither ''mental'' nor ''physical''—there are only mentalistic and physicalistic predicates, which may or may not express real properties' (2003: 9).

which disjunctive physicalism does not deserve the label ‘non-reductive’ (see Bickle 2013; Melnyk 2008; Smart 2007; Fodor 1997; Kim 1992c; Fodor 1974; Putnam 1967; Smart 1959; Feigl 1958). Consider, for instance, Fodor’s assertion that NRPists:

> are required to deny that pain is *identical to* the disjunction of its realisers. The reason they are is that it’s part of their story that the functional property realised, *but not its physical realiser*, is projectible. And the reason they have to say *that* is that *otherwise multiple realisation wouldn’t be an argument against reduction*. What is supposed to make the case for the autonomy (/unreducibility) of functional laws is that there aren’t any laws about the realizer of a functional state *even if there do happen to be laws about the functional state that they realize*. But then functionalists must themselves think that disjunctions–including, notice, the disjunctions that realize bona fide [multiply realised] states–aren’t projectible (1997: 155).

And this is where things get interesting. I hold that disjunctive physicalism is a viable version of reductive physicalism, and that, as we will see, many versions of NRP are variants of it. Thus, I am interested in defending it against the objections of those NRPists, like Fodor, who think that if disjunctive physicalism works, then NRP is in trouble. At the same time, I am interested in criticising the attempts of other NRPists, like Clapp and Antony, to characterise disjunctive physicalism as non-reductive. I shall focus on the latter task in what follows, although some of what I say will also bear on the former.

First, we will look at the objection that disjunctive physicalism is a non-starter because there are no legitimate disjunctive properties (Armstrong 1978). Then we will look at the claim alluded to in the above quote of Fodor’s, that if there are disjunctive properties then they cannot be natural kinds (see Fodor 1974, 1997; Kim 1992c). We will then examine the claim that disjunctive physicalism is a version of NRP because it secures the autonomy of the mental (Clapp 2001; Antony 2003).

No disjunctive properties

David Armstrong has argued that if a predicate is properly disjunctive, then each predicate disjunct will designate a distinct property and this will offend against the principle that a legitimate property is identical in its different instances. As Armstrong states,

> disjunctive properties offend against the principle that a genuine property is identical in its different particulars. Suppose *a* has a property P, but lacks Q while b has Q but lacks P. It seems laughable to conclude from these premises that *a* and *b* are *identical in some respect*. Yet both have the 'property', P or Q (1978: 20; emphasis added).

Consider Armstrong's example: *being a raven or a writing desk*. It does seem implausible to say that all and only the entities in the extension of 'is a raven or a writing desk' are identical in some respect–in some respect, that is, which justifies saying that they are raven-desks (whatever they might be). One way to respond to this is to appeal to other more plausible disjunctive properties which seem to respect Armstrong's *identical in some respect* principle in a way that *being a raven or a writing desk* failed to do; and disjunctivists do just that by appealing to the determinate/determinable relation (henceforth: d/d relation).

Consider the predicate '$F_1 \vee F_2 \vee F_3 \vee \ldots F_n$' such that each disjunct designates a determinate colour property, exhaustively–i.e. every single colour is designated by the disjunctive predicate. '$F_1 \vee F_2 \vee F_3 \vee \ldots F_n$' is necessarily co-extensive with 'is coloured'. Moreover, '$F_1 \vee F_2 \vee F_3 \vee \ldots F_n$' is a properly disjunctive predicate: there is more than one disjunct, each disjunct designates a real property, and each disjunct designates a distinct property. Thus, $F_1 \vee F_2 \vee F_3 \vee \ldots F_n$ is a disjunctive property–one which seems to satisfy Armstrong's *identical in some respect* principle in a way that being a raven or a writing desk does not. Because, every entity in the extension of 'is a raven or a writing desk' will not be a raven-desk (there are no such things), but each entity in the extension of '$F_1 \vee F_2 \vee F_3 \vee \ldots F_n$' will be coloured–necessarily so.

Thus, given the d/d relation, Armstrong's objection fails. And other examples, appealing to other relations along similar lines, are relatively easy to produce. Consider the life/domain relation, and then consider the predicate '$F_1 \vee F_2 \vee F_3 \vee \ldots F_n$' such that each disjunct designates a determinate domain property, exhaustively–i.e. every single domain is designated by the disjunctive predicate. '$F_1 \vee F_2 \vee F_3 \vee \ldots F_n$' is necessarily co-extensive with 'is life' (granting for the moment the completeness of biology).[22]

So, as long as we do not restrict our view to ravens and writing desks, Armstrong's objection can be resisted. However, it is also worth noting that there might seem to be two principles in play here. First: that for every two properties there is a disjunctive property which has those two properties as disjuncts. Second: there are disjunctive properties. Armstrong's objection would work against the first principle, but not, as we have seen, the second.

No Disjunctive Natural Kinds

We will begin by setting up the objection in the terms that were assumed in its original framing. What is a natural kind? In the sense of interest in this debate, F is a natural kind iff F is referred to by an essential law of a complete scientific theory. *Hydrogen*, *Zinc*, and *Sulphur*, we can be reasonably sure are natural kinds in the required sense–they are mentioned in laws which seem essential to one of our most complete theories. *Jade*, *haggis*, and *everything for sale in Honolulu this Sunday* are plausible examples of non-natural kinds. Natural kinds are often said to be 'projectible' (the above quote from Fodor uses the term in this way)–we should read this as meaning simply that the kind in question is *nomic* i.e. figures in natural laws.

The objection: if psychology is a science, then mental predicates designate mental properties which are natural kinds. The disjunctive physical predicates which are co-extensive with the mental predicates do not designate natural kinds (although their disjuncts presumably do); rather they designate a variety of properties which are nomically heterogeneous (Fodor 1974; 1997; Kim 1992). Thus, if psychology is a science, then disjunctive physicalism must fail. Psychology is a science, so disjunctive physicalism fails. This argument can be generalised to cover kinds in other sciences. Recall the above quote

[22] The highest biological taxa is *life*, which, according to the three domain system, divides into three domains of cellular life: *archaea* and *bacteria* are micro-organisms which have no nuclei, and *eukaryota* are any organism constituted by cells which have nuclei and organelles contained within a membrane.

from Fodor about Gresham's law, wampum, cheques and dollar bills.[23] However, given this line of reasoning, it might seem as if proponents of NRP face a dilemma, as follows.

Consider the mental predicate 'M' and the exhaustive disjunctive physical predicate which, in virtue of dependence, it is metaphysically co-extensive with '$P_1 \vee P_2 \vee P_3 \dots \vee P_n$'. The following biconditional holds with necessity: $\forall x\ (Mx \leftrightarrow (P_1x \vee P_2x \vee P_3x \dots \vee P_nx))$. A similar biconditional will hold for every mental predicate and the disjunctive physical predicate that is coextensive with it. If these bridge principles hold, then one might wonder how '$P_1 \vee P_2 \vee P_3 \dots \vee P_n$' cannot be deemed a natural kind term while 'M' is. This invites 'Kim's dilemma'–here's how he puts the point:

> If [being] pain is nomically equivalent to [the disjunction of its realisers], the property claimed [by Fodor and others] to be wildly disjunctive and obviously nonnomic, [then] *why isn't pain itself equally heterogeneous and nonnomic as a kind*? Why isn't pain's relationship to its realisation bases […] analogous to Jade's relationship to jadeite and nephrite? If jade turns out to be nonnomic on account of its dual 'realisations' in distinct microstructures, why doesn't the same fate befall pain? After all, the group of actual and nomologically possible realisations of pain, as they are described by [multiple realisation] enthusiasts with such imagination, is far more motley than the two chemical kinds comprising jade (1992c: 16).

Hence the dilemma: either '$P_1 \vee P_2 \vee P_3 \dots \vee P_n$' designates a natural kind or it does not. Both horns seem to be the death of NRP.

First horn: if the disjunctive predicate designates a natural kind then so does 'M', but then NRP is in trouble because it has been admitted that '$P_1 \vee P_2 \vee P_3 \dots \vee P_n$' designates a natural kind–which directly contradicts the long-standing claims of NRPists that physical realisers of mental properties are far too heterogeneous to form a natural kind. Recall the final part of the Putnam quote from above about how the reductionist might try to 'save

[23] As Fodor stated: 'Gresham's law says something about what will happen in monetary exchanges under certain conditions. I am willing to believe that physics is general in the sense that it implies that […] any event which falls under Gresham's Law […] has a true description in the vocabulary of physics and in virtue of which it falls under the laws of physics. But banal considerations suggest that a description which covers all such events must be wildly disjunctive. Some monetary exchanges involve strings of wampum. Some involve dollar bills. And some involve signing one's name to a cheque. What are the chances that a disjunction of physical predicates which covers all these events [...] expresses a physical natural kind? In particular, what are the chances that such a predicate forms the antecedent or consequent of some proper law of physics?' (1974: 103).

himself by *ad hoc* assumptions (e.g. defining the disjunction of two states to be a single ‘physical-chemical state), but this does not have to be taken seriously’ (1967: 165). The reason it does not have to be taken seriously was supposed to be that such disjunctions would be wildly heterogeneous and couldn’t possibly form a natural kind. But the first horn threatens that very possibility.

Second horn: if ‘$P_1 \vee P_2 \vee P_3 \ldots \vee P_n$’ does not designate a natural kind then neither does ‘M’, but then NRP is in trouble because it has been admitted that mental predicates do not designate natural kinds, and thus psychology is not a science. This would invite the elimination, never mind reduction, of mental kinds.

NRPists might support the contention that ‘M’ but not ‘$P_1 \vee P_2 \vee P_3 \ldots \vee P_n$’ designates a natural kind by appeal to the apparent nomic character of the former in contrast to the apparent non-nomic character of the latter. Note, for instance, Fodor’s rhetorical question: ‘what are the chances that [any disjunctive] predicate forms the antecedent or consequent of some proper law of physics?’ However, another dilemma seems to face NRPists given this line of reasoning too. Consider a law of psychology with the form:

PsychLaw: $\Box\ \forall x\ (M_1x \rightarrow M_2x)$

Here and below ‘$\Box$’ has at least nomic force; ‘M_1x’ and ‘M_2x’ are predicates of psychology. For each psychological predicate, the following bridge principles hold:

Bridge1: $\Box\ \forall x\ (M_1x \leftrightarrow (P_1x \vee P_2x \vee P_3x \ldots \vee P_nx))$

Bridge2: $\Box\ \forall x\ (M_2x \leftrightarrow (P^*_1x \vee P^*_2x \vee P^*_3x \ldots \vee P^*_nx))$

Here, each ‘P’ is a predicate of physics. If both bridges hold then the following physical law is strictly equivalent to *PsychLaw*:

PhysicLaw: $\Box\ \forall x\ ((P_1x \vee P_2x \vee P_3x \ldots \vee P_nx) \rightarrow (P^*_1x \vee P^*_2x \vee P^*_3x \ldots \vee P^*_nx))$

Thus *PhysicLaw* seems to be a law just to the extent that *PsychLaw* is. But in that case, the disjunctive predicate ‘$P_1x \vee P_2x \vee P_3x \ldots \vee P_nx$’ is nomically heterogeneous just to the extent that ‘M_1’ is. Thus another dilemma is faced: either *PhysicLaw* is a law or it is not. If it is a law then so is *PsychLaw*, but then it seems as if all psychological laws can be reduced to physical laws. If *PhysicLaw* is not a law then neither is *PsychLaw*, but then there are no laws of psychology and thus no science of psychology (see Clapp 2001).

Fodor tries to deny that *PhysicLaw* is really a law while affirming that *PsychLaw is* a law by asserting that laws must not be 'gerrymandered'. As he states:

> The way out of [Kim's dilemma] for functionalists is to require that bridge laws be not just nomologically necessary but also that they be *not gerrymandered*. What's wrong with [*PhysicLaw* is that it] *is gerrymandered*. I.e. [*PhysicLaw*] isn't 'independently certified'. I.e. it doesn't occur in any proper ('single level') laws. Since [*PhysicLaw*] isn't independently certified, [*Bridge1* and *2* aren't bridge laws] and (ceteris paribus) [*PsychLaw*] isn't reducible after all. On this account, the constraints on bridge laws are stronger than (in effect, they include) the constraints on proper (single-level) laws (Fodor 1997: 157).

The first thing to note about this is just how strong the requirement is that Fodor is stipulating for any reduction to go through: the relevant predicates of any bridge laws must be predicates used to state the laws of the reducing science. But it is not clear what that requirement has to do with *reduction*. The dialectical usefulness of the condition for Fodor is clear enough: it more or less rules out all reductions except those involving a one-to-one bridging between the laws of the reduced and reducing science. But notice how much this would rule out by considering a textbook case of reduction: classical thermodynamics to statistical thermodynamics. In the above quote from Block and Stalnaker, the short version of the reduction was given: temperature is molecular kinetic energy. But this is not quite right. As we touched on above, and has been much discussed in the philosophy of science literature (see Bickle 2013: § 2.2), temperature *in a gas* is identical to mean molecular energy; temperature *in a solid* is identical to mean maximal molecular energy; temperature *in plasma* is identical to the kinetic energy per particle (in plasmas there are no molecules).[24] Thus, we have a disjunction of predicates which form a microphysical law co-extensive with a macrolaw of classical thermodynamics linked by the relevant bridge laws, in line with the model we developed above. According to Fodor's position, *even this* can't count as reduction simply because the relevant predicates of any bridge laws must be predicates used to state the laws of the reducing science, and the disjunctive law which mentions mean molecular energy in gases, mean maximal molecular energy in solids, and

[24] I put to one side here the more complex question of temperature in a vacuum.

kinetic energy per particle in plasmas is not itself a law of statistical thermodyamics. Thus, Fodor's position is so strong that it rules out the most plausible candidates for reduction that there are. This alone should bring Fodor's position into serious question.

Moreover, our concern here is with ontological reduction, so the issue of what requirements we are to put on laws is somewhat orthogonal. It has been shown that temperature is nothing over and above mean molecular energy in gases, or mean maximal molecular energy in solids, or kinetic energy per particle in plasmas, in the sense that temperature just *is* that disjunctive property. And the case seems to extend to mental properties and their physical realisers. If the mental predicate and the co-extensive disjunctive physical predicate designate the same property, then reductionism in the sense employed herein follows: mental properties are identical to physical properties.

Mental Autonomy

Some have argued that psychology is *autonomous*, and that this is all NRPists need to defend, even allowing that mental kind terms are coextensive with disjunctive physical kind terms, and psychological laws can be reduced, in principle, to physical laws (Antony 2003; see also Clapp 2001: 135–although Clapp puts the point in terms of the non-reducibility, for beings like us, of mental predicates). The word 'autonomy' is often used in connection with discussion of NRP, but it is not always clear what it is supposed to mean. One of its most frequent users, Antony, has this to say about autonomy:

> What matters, fundamentally, is […] whether there is a level of reality beyond the level at which brains are normally studied–whether psychological kinds are 'really there', 'over and above' the already recognised kinds in chemistry, biology and the other established sciences. If this is what is at stake, then it would not matter if brains turned out be the only kinds of things that realise minds in any nomologically possible worlds […] The functional descriptions and the generalisations given in the terms of the psychological categories defined at the functional level would still […] be *autonomous* from the descriptions, generalisations and categories that turned up at the level of the realisers […] Some groupings and regularities are only 'visible' to us at more abstract levels of apprehension, and since many of these groupings and regularities are of enormous importance to us as human beings, we have crucial need for vocabularies that

> capture them. Hence, higher-order vocabularies, and the sciences that embrace them, are, for all practical human purposes, ineliminable (2003: 8, 11; emphasis added).

Two criteria for a kind term to be autonomous are appealed to here: (1) that it be indispensable (except, perhaps, for supreme beings); (2) that it designate a genuine part of our ontology. Moreover, we have ground for believing an autonomous kind term designates a genuine part of our ontology if it is ineliminable. Key to the apparent ineliminability of psychological kinds is their projectibility. This brings us to the key question: how is one supposed to be a disjunctive physicalist, but not a reductionist? The answer, it seems, is that the reductionist will view the nomicity and hence projectibility of psychological kind terms as simply the nomicity of the disjunctive physical kind terms with which they are co-extensive; NRPists partial to disjunctive physicalism, or at least some version of it, will deny this (Antony 2003: 13—15).

Such a denial is not easy to get to grips with. But the basic idea is that mental kind terms are autonomous because they *overlap* rather than *match* each disjunct of disjunctive physical kind terms–this appears to be the only way to justify the claim that mental kind terms are *autonomously projectable*, whilst also being co-extensive with disjunctive physical kind terms. In other words, in order to reject the reductionist's claim that the projectability of psychological kind terms just is the projectability of physical kind terms, the NRPist needs to show that psychological kind terms *carve the world differently*: that mental kind terms designate the overlap among the physical properties designated by the disjunctive physical predicate coextensive with each mental kind term.

Clapp puts the thought in a similar, although distinct way: the disjuncts of a disjunctive predicate overlap iff there is some property *F* such that if a particular is in the extension of any of the disjuncts, then it must instantiate *F* (2001: 126). For example, there is necessarily *overlapping* among the disjuncts of disjunctive predicates which designate determinates of a determinable, and the predicate which designates the determinable picks out all and only that overlap. This is true in the case of the exhaustive disjunctive colour predicate: if an object is designated by, say, 'is red367' then it must instantiate *being a colour*; as well as, of course, *being red367*. In this way, then, one can say that 'is a colour' designates the overlap among each disjunct of the exhaustive disjunctive colour predicate: '$F_1 \vee F_2 \vee F_3 \vee \ldots F_n$'.

Thus, the thought goes, ‘is a colour’ *carves the world differently* compared to ‘$F_1 \vee F_2 \vee F_3 \vee \ldots F_n$’, and will be *autonomously* predictive and explanatory. For example, one could design a colour detection machine to respond only to any and all colours by emitting a sound. Therefore, if those with significant cognitive limits (i.e. humans), are to predict when the machine will sound, and to explain why it does, then they will have to use ‘is a colour’, and the projectability of this predicate is not simply the projectability of a physical predicate–that can’t be the case as the predicates carve at different joints. For mental predicates, an analogous state of affairs will hold, and that is why, it might seem, such predicates are ineliminable and thus autonomous.

So what’s the problem? I’m not convinced that the d/d relation is compatible with NRP (as I will argue in Chapter 3). But independent of that issue, if the proposal works, then I think it does the reductionist’s work for them. Remember, the main obstacle in the way of reduction is the apparent difficulty in discerning any physical property with which to identify mental properties. This difficulty is supposed to be the result of multiple realisation–one can’t find any one way of being physical that matches up with one way of being mental. But notice that the above proposal entails that *all* the realisers of any mental property, M, will *form a resemblance class* such that *M can be identified with that resemblance class*–just as all the ‘realisers’ of being a colour will form a resemblance class. That, so far as I can see, is the Holy Grail of those known for their doctrinaire *naiveté*: mental-physical reductionists.

It will be helpful to retrace the dialectic to finalise this argument. We began by characterising RP as the view that mental properties are identical to physical properties, and NRP as an attempt to deny that view but remain physicalist. NRPists initially denied that mental properties could be identical to physical properties by pointing to actual and possible multiple realisations of mental properties by heterogeneous physical properties. This argument produced an anti-reductionist consensus among physicalists. Then the problems began to appear, the most pressing of which was Kim’s dilemma: a mental property is identical to a disjunction of its realisers, by NRP hypothesis, so two horns bear down on the non-reductionists. First, if the mental property is a natural kind then so is the disjunctive physical property, and reduction follows. Second, if the physical property is not a natural kind, then neither is the mental property, and NRP is left defending a bunch of non-natural mental properties of little interest to science. Some NRPists tried to respond to this dilemma by making the disjunctive option their own, despite the fact that it had previously had been scorned by non-reductionists as an option that ‘does not have to be taken seriously’ (Putnam 1967: 165). The key move towards that end was to redefine NRP as not denying that mental properties are physical properties–which was to be accepted–but instead as assuring that psychology was autonomous from physical science.

But note, this argument can go in more than one direction. One could argue that disjunctive physicalism has shown that NRP is compatible with the identity of mental and physical properties, or one could argue that disjunctive physicalism has shown that RP is compatible with multiple realisation and the autonomy of psychology. Which way we go here is largely determined by how we are to settle the question of what ‘reduction’ means–a question we have already dealt with: reduction in the present context is entailed by mental-physical property identity. And what do we say to someone who denies this? ‘Reduction’ is a term of art, so people may use it as they please, but one condition on its use that ought to be respected is dialectical relevance–the extent to which any new use of a term is relevant to previous uses of the term in a given dialectic. As we have established, the dialectic we are engaged in here is the dialectic that began (let us pick a precise date for dramatic effect) on May 15th 1959 in Washington Square New York, when Putnam first spoke of minds and machines. In *that* dialectic, which continues to this day, ‘reduction’ is taken to refer to a mental-physical property identity thesis. Thus, given the terms of that dialectic at least, disjunctive physicalism, if it works at all, works as a demonstration of the compatibility of reduction with multiple realisation and the autonomy of psychology.[25]

1.6 Conclusion

We began by considering the origin of the dispute between reductive and non-reductive physicalists: whether mental properties are identical to physical properties. NRPists denied this by appealing to the multiple realisation of mental properties by physical properties–the founding idea of classic NRP. We also considered Donald Davidson’s anomalous monism, which, we saw, fails to meet the requirements of physicalism. Then came disjunctive

[25] One might object at this point: disjunctive physicalism can’t licence identifying mental with physical properties because Cartesian worlds are metaphysically possible, and so, pain, say, can’t be identical with a disjunction of physical realisers – Cartesian pains are certainly not physical. This objection raises difficult questions about the nature of physicalism and the possibility of Cartesian worlds. But I think that the best response on behalf of the disjunctive reductionist is to allow for Cartesian worlds which have mental properties, but to deny that those mental properties are identical to the mental properties in this world.

physicalism, advertised by its proponents as a variety of NRP which could meet reductionists' most powerful objection: Kim's dilemma. However, it was demonstrated that disjunctive physicalism proves too much: if it works as view, then it works merely to show that reductive physicalism is compatible with both the realisers of mental properties forming a natural kind, and the autonomy of psychology. It is difficult to imagine a more resounding triumph for the philosophy of mind's Old Left.

This leaves classic NRP, or more contemporary versions which are not disjunctive physicalism, guarding the non-reductive fort. In our next two chapters, we will examine a major problem faced by NRP: the overdetermination problem, and the efforts to try and solve that problem of a new wave of NRPists that I label 'immanentists'. As we shall see, however, there are significant problems both with the overdetermination problem itself, and with the solutions offered by immanentist NRPists.

Chapter 2
The Super-Overdetermination Problem

Chapter Summary

In the last chapter, we examined the debate between RP and NRP, characterising both views and clarifying the point at issue between them: whether mental properties are physical properties. In this chapter we turn now to consider *the* major problem that NRP is supposed to face.

Non-reductive physicalism is thought to encounter an overdetermination problem. But the traditional articulation of the purported problem is flawed because it fails to articulate a *genuine* problem. I examine the original intuition on which the traditional articulation is based, and show how to articulate a genuine problem for non-reductive physicalism which answers to the original intuition. I call my articulation the 'super-overdetermination problem', and demonstrate that it is best viewed as a special case of the more general problem of explaining the holding of the relation between the mental and the physical–the mind-body problem, in essence. I then contrast my super-overdetermination articulation with four disambiguations of the traditional articulation, showing why each is not a genuine problem for non-reductive physicalism. This is followed by a detailed critique of the most famous version of the traditional articulation: Jaegwon Kim's exclusion argument. Penultimately, I discuss how RP faces a problem which is similar to the genuine problem NRP faces. We finish by considering an explanation for the near five decade persistence of the traditional articulation of the overdetermination problem.

The human body is assumed to be as complete a causal system as is a gasoline engine.
- Norman Malcolm[26]

2.1 Articulating Malcolm's intuition

There is supposed to be a problem of overdetermination faced by non-reductive physicalism. There is a problem with this problem, however. The consensus on how to frame the problem (broadly) is not matched by a similar degree of consensus on whether this framing really frames a *problem* at all. As we will see, this can be resolved. Once the issue is properly outlined, it can be shown that there is a genuine problem of overdetermination faced by non-reductive physicalism, it's just that since its introduction into the discourse nearly half a century ago, everyone has failed to articulate it properly.

To see this, it will be useful to begin by examining the original intuition upon which the discourse is ultimately based, and which, in general form, still permeates the literature. In 1968, Norman Malcolm argued that there was an incompatibility between two kinds of explanation of bodily movement: 'mechanistic' versus 'purposive'. Mechanistic explanations are the kind of explanation you might give of how an engine works. For example:

In order to turn gasoline into vehicular motion, standard internal combustion engines work in a four stroke combustion cycle: intake; compression; combustion; exhaust. As the piston drops, the piston cylinder is filled with fuel and air through the intake valve. The momentum of the piston brings it back up thereby compressing the fuel-air mixture. The spark plug then sparks, igniting the gasoline, which explodes–pushing the piston back down. As the piston returns from the bottom of its stroke, it pushes the exhaust out of the cylinder via the exhaust valve. The process repeats.

This contrasts with purposive explanations such as: the piston *wants* to go up and down, so the piston goes up and down.

Clearly, in the case of combustion engines, mechanistic explanations are to be preferred to purposive ones. Perhaps the purposive explanation would be better than nothing–one might

[26] Malcolm (1968: 45).

give such an explanation to a small child–but once in possession of the mechanistic explanation, any purposive explanation is excluded–it loses the 'competition'. Malcolm's worry is that something similar could hold for the human body: that mechanistic, neurophysiological explanation of bodily movement could compete with psychological explanation, and that we might have to conclude that one type of explanation loses out to the other. That's the original intuition; the place from whence the overdetermination problem came.

Call this 'Malcolm's intuition'–the intuition that there appears to be some sort of problematic competition between psychological and neurophysiological explanations of bodily movement.

It's in the nature of intuitions that mark the first step in a philosophical investigation that they be indefinite, leaving room for divergence in subsequent attempts to articulate more robust versions of the original intuitive concern. It is my contention that from Malcolm onwards, every attempt has missed the mark.

In order to see how best to articulate Malcolm's intuition, let's begin, as charity demands, by granting that Malcolm was actually on to something. If he was, then I think two conditions must be met if the intuition is to be properly articulated:

(AC1) a *genuine* problem is articulated. A genuine problem for a theory requires the theory to change in some way. The theory can either be appropriately amended without giving up an essential tenet, in which case the problem is solved, or the theory can't be so amended, in which case it must be abandoned.

(AC2) The articulation does justice to Malcolm's original intuition of causal competition. Doing justice to an intuition is a matter of giving a plausible account of what the intuition was about. Whether or not an articulation does this is best judged by simply comparing the intuition to the articulation.

Call AC1 and AC2 the 'articulation conditions'. Although the second condition has been met by other commentators, the first condition has not. As I will show, my articulation, in the form of the super-overdetermination problem, satisfies both.

The argument in summary:

In Section 2.2, I outline the theory which is the primary target of the overdetermination problem: non-reductive physicalism (henceforth: NRP).

In Section 2.3, I lay out what I take to be the best articulation of Malcolm's intuition given that NRP is the target, what I call the 'super-overdetermination problem', and I show how it meets AC1 and AC2.

Because of the close similarities between the super-overdetermination and superdupervenience problems, in Section 2.4 I discuss the issue of whether or not they are distinct problems at all. I argue that because the range of possible solutions to each problem is different, the problems are not identical. I suggest that the super-overdetermination problem is best viewed as a special case of the more general superdupervenience problem, and that this has important consequences for the assessment of the viability of NRP: if NRP can't solve the former, then it can't solve the latter.

In Section 2.5, I explain the difference between super-overdetermination and four other kinds of overdetermination: 'flukey', 'incoherent', 'incomplete' and 'complete'. These four kinds of overdetermination are often not clearly distinguished in the debate, but one or another of them has erroneously been deemed to be *the* problematic kind of overdetermination faced by NRP. I show that none of them are genuine problems for NRP, and so fail AC1.

In Section 2.6, I examine the best-known articulation of Malcolm's intuition, as expressed in Jaegwon Kim's exclusion and supervenience arguments, which are descendants of Malcolm's own articulation. I show how Kim's articulation fails AC1, and I discuss the role played in his arguments by the exclusion principle, productive causation, and supervenience.

In Section 2.7, I demonstrate that reductive physicalism faces a problem similar to that faced by NRP, namely 'epistemic-overdetermination', although there remain important differences between the problem faced by NRP and RP.

In Section 2.8, I propose an explanation of the persistence of the traditional articulation of the overdetermination problem that NRP is supposed to face.

In Section 2.9, I conclude.

2.2 The target: non-reductive physicalism

A rhetorician could make good use of the truth: ‘there are no problems’. It is true because, in the sense of ‘problem’ of interest to us, there are no problems *simpliciter*; there are only problems *for* this or that theory. Thus, if we are to properly articulate Malcolm’s intuition into a genuine problem in line with the first articulation condition, then we need to have a clear understanding of the theory we are trying to articulate a problem for. According to the standard story, that theory is NRP. Recollect that the core tenets of NRP are the following necessary conditions:

The nothing above tenet: Mental properties depend on physical properties.
The distinctness tenet: Mental properties ≠ physical properties.
The causation tenet: Mental properties are causally relevant for physical properties.

We have already discussed these tenets in Chapter 1 (Section 1.3), but for presents purposes it will be helpful to say a little more about the first and the third tenet.

With regards to the dependence claim, there is widespread agreement that NRP is viable only if the holding of dependence can be explained in a materialist way (Schiffer 1987; Kim 1990a; Heil 1992; Horgan 1993; Wilson 1999; McLaughlin and Bennett 2014). There is not the same widespread agreement on what would count as explaining the holding of dependence in a materialist way (see, for instance, the debate between Horgan 1993 and Wilson 1999). This is a genuine problem for NRP that has driven much of the effort towards developing more sophisticated versions of the view. Let us press a popular term into service, and call the problem of providing a materialist explanation of the holding of dependence the ‘superdupervenience problem’.[27]

[27] ‘Superdupervenience’ was coined by Lycan (1986: 92n19) and popularised in the philosophy of mind by Horgan (1993).

It is not a straightforward matter to spell out exactly why the superdupervenience problem is a genuine problem for NRP. Often, it is framed in terms of showing that NRP is inconsistent with emergentism. As we saw in Chapter 1, NRP is supposed to be a third way between reductive physicalism on the one hand, and emergentism on the other, where emergentism is understood as a non-physicalist view which holds that the mental depends on the physical in virtue of non-physical fundamental laws of emergence. In order to rule out that NRP is not just emergentism in disguise, the thought goes, NRPists need to explain why their appeal to mental-physical dependence is not just a non-physical law of emergence. Crucially, if they fail to do so, then they fail to show how their view is a physicalist view, rather than just emergentism in disguise.

I think there is more to the problem than this, but this rough sketch will do for present purposes until we deal with the problem in detail in Chapter 3, where I give an account of what's really at issue in the superdupervenience problem, and provide an explanation of the three explanatory criteria that I think NRPists must meet in order to solve the superdupervenience problem. I will very briefly outline those criteria now.

The superdupervenience problem can be solved by NRP only if it can be shown that there is a relation, or set of relations, R, such that:

(i) R entails dependence;
(ii) R is internal;
(iii) R holds between the mental and the physical.

Let us call these the 'three criteria'. The first criterion is required because if R doesn't entail dependence then it can't be a sufficient explanation of why dependence holds, which is what is required if NRPists are to rule out that their view is not physicalist.

The second is more complicated, but here's the basic idea. Think back to the regress problem faced by those trying to explain the holding of Nagelian bridge laws. That problem was an instance of a more general problem of explaining the holding of trans-ordinal relations consistently with physicalism. A trans-ordinal relation is just a relation that holds between the mental and the physical (or other domains).[28] But the problem with trans-ordinal relations, from a physicalist perspective, is that they are, well, trans-ordinal–they stand *between* the mental and the physical. This makes it hard to explain in physical terms why they hold. We saw one example of how hard this is with the regress problem faced by trans-ordinal Nagelian bridge laws. As I will discuss in Chapter 3 (see Section 3.2), NRP also faces a version of this regress problem, whereby they are required to explain the holding of mental-physical dependence by appeal to R, but R can't just be a fundamental non-physical relation, so NRPists have to rule that out. NRPists can't do this, of course, just by appealing to a further R*, as this begins the regress. The best way out of this problem for NRP, I suggest, is to show that R is internal. Minimally, a relation, R, is internal to x and y iff, necessarily if x and y exist then xRy. The most obvious way to show that R is internal is to show that the natures of x and y entail that they stand in R. For example, it could be the case that the existence of Jack (x) and the existence of Jill (y) necessitates that Jill stands to Jack in the *taller than* relation (R), and we could show this by appeal to Jill's being 6 feet tall, and Jack's being five feet 8 inches tall.

I consider in more detail how this would work in the mental-physical case in Chapter 3 (esp. Section 3.2), but the basic idea is that R's holding can be explained by appealing to R's relata–rather than R's holding being explained by appeal to R*–and in this way the regress problem can be avoided.

For the third criterion, it is clear that even if an R which entails dependence and is internal can be posited, this can still fall far short of explaining why dependence holds. For example, it could be contended that the mental stands to the physical in the relation: *is a higher biological taxa than*. But the mental is clearly not a higher biological taxa than the physical, so this demonstrates that R also needs to be shown to actually hold between the mental and the physical.

NRPists have put forward plausible proposals for how to explain the holding of the dependence relation from when the view first came to dominance in the 1960s. Usually, R has been labelled 'realisation' and a theory offered of what 'realisation' refers to alongside a defence of the claim that the mental *is* realised by the physical. However, whatever relation NRPists have posited, meeting these criteria has proved to be a particularly difficult challenge (although the efforts of NRPists have not always been aimed at these

[28] The term 'trans-ordinal' was coined by C. D. Broad (1925: 52).

criteria in particular). This is important because, as we will see, the overdetermination problem for NRP is a special case of the superdupervenience problem, and so solving the former will be just as difficult as solving the latter.

I will say much more about the three explanatory criteria in Chapter 3 (Section 3.2). Let us now finalise our characterisation of the target: NRP.

It will be helpful to give a reading to NRP's third tenet which will allow us to simplify the dialectic. Consider the contrast we encountered in Chapter 1 between property and event dualism (Section 1.4). If you believe that distinctness only holds for mental properties and that mental events are identical to physical events, then questions of overdetermination are questions about two apparently competing properties (*à la* Davidson). Alternatively, if you believe that distinctness holds not only for mental properties but also that mental events are distinct from physical events, then questions of overdetermination also involve questions about two apparently competing events. Given the traditional way of articulating the overdetermination problem, it might seem to matter whether one is an 'event monist and property dualist' versus a 'property and event dualist'. One could ask, for example: is property causal relevance overdetermination more or less problematic than event overdetermination? If you wish to defend the traditional articulation, then I leave that question with you. I wish to show the traditional articulation to be wrongheaded, for reasons which trump any answer to this question. Thus, for ease of dialectical exposition alone, I will be assuming a version of NRP that is property and event dualist.

This gives us the basic version of NRP that will be our target:

The nothing above tenet: Mental properties depend on physical properties, and mental events depend on physical events.

The distinctness tenet: Mental properties ≠ physical properties, and mental events ≠ physical events.

The causation tenet: Mental events cause physical events, and mental properties are causally relevant for physical properties.

Of course, proponents of NRP have much more to say about their view than this basic model. But that it is so basic gives broad scope to my criticisms of the efforts to articulate an overdetermination problem *for NRP*. Moreover, as we shall now see, it is in trying to move from this basic model towards more substantial theses that NRPists run into the genuine problem of overdetermination.

2.3 The super-overdetermination problem

I think that the *real* overdetermination problem faced by NRP has been missed so often, and the terminology in the general debate has become so muddled, that a new name which unequivocally picks out the real issue will be helpful; hence: the *super*-overdetermination problem.

In order to show that the super-overdetermination articulation is *the* way to articulate Malcolm's intuition, I will first show that it is *a* way to articulate Malcolm's intuition, before, in Section 4, ruling out the alternative ways.

So what is super-overdetermination? In the most general sense, super-overdetermination of a physical effect by a mental and a physical cause occurs iff the causes are:

1) distinct from each other;
2) each sufficient for the effect;[29]
3) unmediated with respect to the effect;[30]
4) and the particular necessitation relation they stand in to each other is metaphysically brute.

[29] The sufficiency I have in mind here is not sufficiency *simpliciter*, but rather sufficiency given a context being fixed first. For example, given the presence of oxygen plus other conditions and the relevant natural laws, the spark can be a sufficient cause of the fire. In contrast, the big bang was a sufficient cause *simpliciter* (let us assume) of the immediately following state of the universe.

[30] By 'unmediated' I mean that there are no other events entering into the causal chain between the causes and the effect. When framing the problem of overdetermination purportedly faced by NRP, people speak of 'synchronous' or 'immediate' causes instead of 'unmediated'. I prefer the 'unmediated' condition to either of those because synchronicity of causes allows for awkward cases where there's mediation: e.g. M is synchronous with P, M is sufficient for M_1 and P is sufficient for M_2, M_1 and M_2 are in turn both sufficient for P*, although M_1 and M_2 aren't synchronous. Also, opting for the 'immediate' condition is ambiguous between spatial and temporal immediacy, neither of which seems required as part of the definition. Because, demanding temporally or spatially immediate causes might, at least on some views, be too strong – ruling out that there can be super-overdetermination when the causes are either temporally or spatially distant from the effect. Whether or not such cases are possible is controversial, so it would seem unwise to rule them out in the very definition of super-overdetermination. What really matters, for our purposes, is that the causes are *unmediated*.

By ‘necessitation relation’ I mean any non-causal relation which holds synchronously and has at least some degree of modal force. Of course, with NRP, the particular necessitation relation in question is dependence. As I use the terms, a relation can be *brute* in an epistemic or metaphysical sense. A relation is metaphysically brute if it’s holding is not necessitated by anything else (‘metaphysically brute’ is synonymous with how many use ‘fundamental’). A relation is epistemically brute if there is no true, informative explanation for why we should believe that the relation holds.

In order to rule out that the dependence relation posited by NRP to hold between mental and physical events is brute, NRPists must meet the three criteria as applied to events; namely, show that there is a relation (or set of relations), R, such that:

(i) R entails dependence;
(ii) R is internal;
(iii) R holds between mental and physical events.

The super-overdetermination *problem* for NRP is simply that if a theory entails that super-overdetermination occurs, then that theory is not physicalist. The theory would not be physicalist because if physicalism is true, then the dependence relation which holds between the mental and physical is not fundamental, i.e. in the sense of not being necessitated by anything else (what I also call being metaphysically brute). No physicalist can countenance non-physical fundamentals–they must be ruled out. As the mantra goes, everything is physical or depends on the physical.

Thus, physicalists need to rule out super-overdetermination, to fail to do so is to fail to be physicalist. *That* is the super-overdetermination problem. See how it meets the articulation conditions:

(AC1): it’s certainly a *problem* for NRPists because physicalists can’t allow super-overdetermination.

(AC2): It does justice to the intuitive notion of problematic causal ‘competition’ that Malcolm began with and which still permeates the literature. Consider the metaphor: in a competition typically there has to be a winner, or at least an explanation, in light of the rules, of why there can be more than one winner, or none at all. Now consider super-overdeterminers: they're each distinct, sufficient and unmediated causes, so we might be tempted to say, when looking at them individually, '*that* cause is *the* cause' (i.e. *the* winner). But when looking at the situation overall, it appears that we can't say that either one or the other cause is *the* cause (i.e. *the* winner) because of the dependence relation, in virtue of which the two causes are 'up against each other'. We might then be tempted simply to say that *both* can be considered the causes (i.e. *both* are winners). But we're physicalists, so the rules tell us that we can't just help ourselves to a brute dependence relation here, so in the absence of an explanation of the dependence we can't simply accept that both causes are winners. Similarly, saying that *neither* cause is *the* cause (i.e. it's a *draw*) would only work within the rules–i.e. if we have a physicalist explanation of why there is a 'draw', an explanation of why the causes stand in the dependence relation.[31]

There are three objections to this articulation that it makes sense to get out of the way quickly.

First objection: 'but no NRPist thinks that mental causes brutely depend on physical causes, so how is super-overdetermination a problem for NRP?' But this misses my point: the issue here is about how to frame the debate correctly. The question I'm trying to answer is: *what's the goal that NRPists have to reach if they are to solve the problem of overdetermination faced by NRP?* My answer is that the goal is to solve the super-overdetermination problem, and *not* any other purported problem of overdetermination articulated in the traditional way. (And I have begun that answer in this section by showing that the super-overdetermination problem is at least one way of articulating Malcolm's intuition.) Of course, NRPists have candidate explanations of why the mental depends on the physical–that is not in dispute. The claim I am maintaining is that if NRPists are to genuinely solve the problem of overdetermination faced by NRP, then they can do so only if they explain, in the sense I employ herein, the dependence relation between mental and physical overdetermining causes – i.e. by genuinely solving the super-overdetermination problem.

[31] This discussion of how super-overdetermination meets AC2 also provides a response to those who have based part of their argument in defence of NRP by reflecting on the various senses of 'compete' and concluding that if A depends on B, then A can't *compete* with B in any of the senses of 'compete' that have been elucidated (e.g. Garrett 1998). My discussion of how super-overdetermination meets AC2 demonstrates the *problematic* sense in which mental and physical causes 'compete' in the NRP model: it seems like we can't say either cause 'wins', nor that both or none 'win' – *in the absence of an explanation of the dependence relation*.

Second objection: ‘but the problem of overdetermination is supposed to be an *overdetermination* problem, and the super-overdetermination problem is not that sort of problem.’ In many ways, this objection expresses the confusion that has run through the NRP-overdetermination debate since it began. When people say that the problem of overdetermination is supposed to be an ‘*overdetermination* problem’ they typically have in mind the kind of overdetermination that is often illustrated with examples of firing squads and assassins and such like. As I will argue below, this is precisely the *wrong* way to think about the problem of overdetermination faced by NRP. Moreover, this second objection is just false: ‘overdetermination’ has a well-known general sense: A and B overdetermine C if A≠B, and A and B are both sufficient for C. This sense is often remarked upon in the literature. Here’s Simon Blackburn:

> An event is overdetermined if there exist more than one antecedent event, any of which would be a sufficient condition for the event occurring (2008: 229).

This general sense of ‘overdetermination’ covers different varieties of overdetermination. For example, there’s sequential overdetermination: A at t_1 being sufficient for B at t_2 which is sufficient for C at t_3. In a deterministic universe, all events are sequentially overdetermined. There are also varieties of non-sequential overdetermination. Of particular interest to us, of course, is non-sequential unmediated overdetermination.

Thus, the second objection is misplaced, the super-overdetermination problem uses ‘overdetermination’ in a perfectly acceptable sense. And if the objection is simply meant to push the thought that the overdetermination problem faced by NRP can only be articulated in terms of a particular variety of overdetermination that the objector has in mind, then this is exactly the claim that I am contesting here. And I contest it in two ways, first by showing that there is *a* way of articulating an overdetermination problem for NRP in terms of super-overdetermination, which I have just done. Below I show that all the other ways of trying to articulate the problem in terms of other kinds of overdetermination fail the first articulation condition, this leaves my own articulation as the only version of the problem left standing.

Third objection: ‘what about causation – is not the problem of overdetermination faced by NRP supposed to be a *causal* problem?’ If by ‘*causal* problem’ it is meant ‘a problem which is entirely or primarily a problem about the nature of causation’ then I disagree. The original intuition of the appearance of some sort of causal 'competition' does not entail that, as the above discussion demonstrates. Much of the debate around the problem of overdetermination NRP is supposed to face has viewed it as an issue resting on the resolving of disputes about the nature of the causal relation (e.g. Baker 1993; Horgan 1997; Kim 2007a; Loewer 2007; Maslen, Horgan & Daly 2009; Zhong 2011; Christensen & Kallestrup 2012; Papineau 2013; Menzies 2013);[32] but if the real problem is super-overdetermination, then this causation focus looks like a red herring. We should have been going after an explanation of the dependence relation all along.[33]

This is not to say that causation does not enter into the picture at all, however. Analyses of causation may be relevant, but in a manner quite distinct from that assumed by most discussions in the literature. The distinct way that causation may be relevant is that it might play a role in determining one's choice of explanation of the dependence relation. Moreover, as we will now see, in considering how the explanation of the dependence relation might be constrained by causal theory, the particular way in which the super-overdetermination problem is a special case of the superdupervenience problem is revealed.

2.4 Super-overdetermination and superdupervenience

The question: 'what is the relationship between the super-overdetermination problem and the superdupervenience problem?'

The *wrong* answer: the super-overdetermination problem *is* the superdupervenience problem. Why this might *seem* like the right answer: NRP states that there is mental-physical dependence, but unless the holding of that relation can be explained, then NRP doesn't qualify as a physicalist position–that's the superdupervenience problem. And it's

[32] It is difficult to find a contributor to the overdetermination debate who does *not* appeal to some analysis or other of causation in developing their articulation of the problem. This list includes some indicative examples.

[33] The super-overdetermination articulation is somewhat ironic given the widely held view that the necessitation of the mental by the physical is essential for there to be any mental causes of physical effects (e.g. Fodor 1987: 42; Kim 1998: 40). One way to put the thought is that if the mental causes of physical effects don't supervene on the physical causes of those same physical effects, then it would allow for mental and physical causes to overdetermine physical effects in the classic sense where two *independent* events are sufficient unmediated causes of an effect, thus undermining the causal status of the mental cause because the effect could occur without it. Kim has also seen a 'double-edged' role for dependence in relation to mental causation, although of a different sort to the super-overdetermination problem (see Kim 1998: 46). See Section 2.6 for more on this last point.

very hard to explain the holding of the dependence relation– the matter is something of a mystery–so NRP faces a particularly difficult problem. Hence, until a form of NRP can be produced which avoids the superdupervenience problem, then NRP is at best incomplete, and under suspicion of being false. The super-overdetermination problem differs from the superdupervenience problem only in that it's framed in terms of *causes*–but that doesn't add anything to the *problem*, it's all about explaining the dependence relation.

The *right* answer: problems are theoretical entities whose identity conditions include the solution space–the range of possible solutions to the problems. What appear to be two problems are identical only if their solution spaces are identical. What the solution space seems to be changes over time, as theories and understanding develop, so there can appear to be two problems at t_1, before we realise at t_2 that there was really only one problem (or vice versa). Thus, if problem A and problem B appear to have different solution spaces at t, then we should judge, at t, that $A \neq B$.

The easiest way to see the difference between the solution space for the superdupervenience problem and the super-overdetermination problem is to notice that the latter essentially mentions causes, but when specifying the former causes aren't even mentioned. If it were to be discovered, indeed, that there are no mental causes at all, ever, anywhere, then there would be no super-overdetermination problem, but there would still be a superdupervenience problem. Thus, the solution space for the superdupervenience problem looks to be distinct from the solution space for the super-overdetermination problem, and so we should conclude that the appearance of *two* problems is correct.

Moreover, the fact that the super-overdetermination problem essentially mentions causes seems to bring extra baggage with it–namely the relationship between one's causal theory and one's theory of the dependence relation between mental and physical causes. In particular, one's causal theory might include commitments that constrain what options are available for explaining the mental-physical dependence that holds between causes. There are numerous examples of philosophers' commitments regarding causation affecting their theories about the relationship between mental and physical causes. Just think of the issues Descartes had with getting his model of the relationship between mental and physical causes to cohere with what some thought a plausible causal theory would require. Or recall Davidson's anomalous monism, to pick another example, which included the commitment that one event causes another only if there is a strict law under which both events are subsumed, and that only physical events fall under strict laws. Davidson used this claim to help deduce that mental events are identical to physical events.

We need not concern ourselves with the virtues and vices of anomalous monism or Cartesian dualism. Instead, the lesson to take from this is that one's theory of causation can have a significant impact on what one says about the relationship between mental and physical causes. If that's true, then it would seem unwise to rule out that one's theory of causation could constrain the space of possible explanations of the mental-physical *dependence* relation that NRPists assert holds between causes.

Of course, all this leaves open the question of just *how* distinct the two problems are. It seems probable that the best candidate for solving both problems will be the same, i.e. the same relation to explain the dependence between both mental and physical causes and the mental and physical more generally. In that case we should conclude that as the problems are so similar in other ways as well, their having identical solution spaces entails that the super-overdetermination problem is simply a special case of the superdupervenience problem.

If the problems do so collapse, then this has an important consequence for NRP: if NRP can't solve the super-overdetermination problem, then it can't solve the superdupervenience problem. Because if the super-overdetermination problem is a special case of the superdupervenience problem, then the superdupervenience problem can be solved only if the super-overdetermination problem can. Thus, two birds with one stone: we can use the super-overdetermination problem as a test for whether or not NRP can solve both problems, and so gain a very clear understanding of the viability of the view.

The *non*-collapse alternative is that it could be discovered that the dependence between mental and physical *causes* has to be explained in one way–by appeal to a relation R_1, say; and, the dependence between *the* mental and *the* physical (specified in terms of mental and physical *properties*, say) has to be explained another way–by appeal to a relation R_2. If R_1 and R_2 differ significantly (e.g. R_1 is not just a version of R_2), then the superdupervenience and super-overdetermination problems would also differ significantly, in virtue of the significant difference in their solution spaces; and thus the latter would not simply be a special case of the former.[34]

I think the non-collapse alternative is a long-shot. This is because any attempt to explain the dependence between *the* mental and *the* physical will most likely proceed by explaining the dependence between mental and physical properties; and, it is difficult to see how one would explain the dependence between mental and physical events without simply explaining the dependence between the relevant event properties. This gives good grounds to expect the collapse over the non-collapse outcome.

The 'special case' characterisation of the relationship between the super-overdetermination and superdupervenience problems also makes sense of a common refrain in the literature, namely that the problem of mental causation and the mind-body problem are deeply interconnected. Kim, for example, talks of the two problems as intertwined 'world-knots' (2005: 7). In doing so, he is borrowing a famous metaphor of Schopenhauer's that many others have borrowed. The super-overdetermination problem is NRP's version of the problem of mental causation (I contend), and the superdupervenience problem is NRP's version of the mind-body problem. Given the 'special case' framing, the intuition that the problems are intertwined is accounted for, because constituting both 'world-knots' is the same piece of 'rope': a mental-physical dependence relation in need of explanation. The intuition that they are distinct is also accounted for, because although the dependence relation is at the heart of the problem in each case–it's the same piece of rope–one dependence relation is a special case of the latter–hence the two knots to untie.

I have explained why I hold that there's *a* problem of super-overdetermination faced by NRP, and the way it differs from the superdupervenience problem. I also hold that super-overdetermination is *the* problem of overdetermination faced by NRP–we will now see why.

[34] Except in the rather uninteresting sense that the superdupervenience problem would be a more general 'necessitation relation requiring explanation' problem, and the super-overdetermination problem a more specific 'necessitation relation requiring explanation' problem.

2.5 Flukey, incoherent, incomplete, and complete overdetermination

The main thought often driving the traditional articulations of Malcom's intuition is that *the problem* for NRP is that causes must *make a difference* to their effects, but NRP doesn't accommodate mental causes making a difference to physical effects. Because, according to NRP, every time there's a physical effect that is supposed to have a mental cause, it also has a physical cause which 'already' does 'all the causal work'. In this way, the thinking goes, NRP undermines the status of mental events as *causes* of physical effects. Often, this line of thought has been expressed counterfactually by saying something like: 'according to NRP, if the mental "cause" of the physical effect had not occurred, then the physical effect would have occurred anyway–because it has a physical cause; so how can this mental "cause" be considered a *genuine* cause? It does not make a difference!'

The traditional arguments also invariably include a 'no overdetermination' condition as a key premise. As we will see, though, there are significant issues with clarifying and justifying the condition. Indeed, as I will argue, this is an important part of the problem with the traditional ways of articulating Malcolm's intuition, and helps to explain why they fail to articulate a *genuine* problem *for NRP*–and so fail AC1.[35] So, given that the super-overdetermination articulation meets AC1 and AC2, and the traditional articulations fail AC1, we should conclude that the super-overdetermination problem is *the* problem of overdetermination faced by NRP.

To see this, it is necessary to distinguish *super* from four other kinds of overdetermination that people apparently refer to when they talk of *the* 'overdetermination' or 'exclusion' problem in the philosophy of mind. I call those four kinds: (i) flukey; (ii) incoherent; (iii) incomplete; and (iv) complete. The literature fails to draw clear enough distinctions between them. Different commentators appear to have either (i), (ii), (iii) or (iv) in mind when trying to articulate Malcolm's intuition, and it is not always obvious which. Moreover, some commentators slide between talk which suggests either (i), (ii), (iii), or (iv), sometimes across articles, sometimes within the same article. (Specific examples will be provided below.) This can be confusing, but in the end we needn't worry too much about determining what others in fact mean, or should mean, because once we distinguish

[35] I grant that the traditional articulations meet AC2.

and examine (i), (ii), (iii), and (iv) we will see that they all fail the first articulation condition.

Let's consider (i), (ii), (iii) and (iv) in turn.

Flukey-overdetermination of a physical effect by a mental and a physical cause occurs iff those causes are:

1) distinct from each other;
2) each sufficient for the effect;
3) unmediated with respect to the effect;
4) independent (i.e. not standing in a necessitation relation to each other).[36]

The firing squad or two assassins case is the standard example: two shooters, shooting simultaneously, with each shot a sufficient cause of the death of the victim.[37] Unfortunately, that NRP needs to avoid flukey-overdetermination has too often been taken to be *the* overdetermination problem faced by NRP, when it is anything but. Thomas Kroedel, for example, suggests that:

> the rationale behind […] the exclusion problem is that cases where behavioural events have simultaneous mental and physical causes would be similar to prototypical cases of overdetermination such as deaths by firing squads […] in firing squad cases the two overdetermining events *independently* bring about the effect (2013: 11; my emphasis.[38] For similar comments, see Kroedel 2008: 128-129).

[36] I use 'necessitation relation' in the same way here as I did in Section 3: 'any non-causal relation which always holds synchronously and has at least some degree of modal force.'

[37] For simplicity, let us assume that the firing squad or two assassins case we are considering is one where the shooters *don't* have a temporally near joint cause (e.g. an order to shoot, or bounty, or such like), and just shoot simultaneously of their own accord. Also, like most of the examples we will discuss, it seems plausible that the two causes here, the two shootings, *are* mediated with respect to the effect – the death. However, to give examples which are plausibly *not* mediated would require greater complexity for no great expository advantage. It is much easier simply to grant that in the more intuitive and dialectically relevant examples the causes are unmediated.

[38] Note: Kroedel considers a version of the problem faced by an avowedly non-physicalist dualism of the sort most famously espoused by David Chalmers (1996), where mental properties depend on physical properties with only nomic necessity. I agree that this is a non-physicalist position, although others might not (e.g. Kim 2005: 49).

But flukey-overdetermination is not a genuine problem for NRP, and so fails the first articulation condition. Because flukey-overdetermination is simply not relevant to NRP, such cases are not what should worry us when considering NRP. Instead, as we have seen, it is the fact that the mental *isn't* independent of the physical that generates a problem: namely the problem of explaining the holding of the particular necessitation relation that mental and physical causes stand in to one another. And notice that overdetermined deaths by firing squads are *explainable*: the shooters just so happened to shoot at exactly the same moment, the bullets took trajectories such that they hit simultaneously and such that either impact would have been sufficient to kill the victim. Importantly, it is the *independence* of the causes which makes the flukey overdetermination explainable: add in enough causal chains to a universe and overdetermination of this sort is bound to happen at some point.

The literature contains many commentators, like Kroedel, who get matters the wrong way round by thinking that NRPists can respond to overdetermination worries in significant part by *appealing* to the fact that the mental is necessitated by the physical *without explaining the holding of that necessitation*. Indeed, *all* of those who have argued that NRP can rid itself of worries about overdetermination without addressing the issue of explaining the necessitation between the two causes make something like this error: the real problem is simply left unaddressed, obscured behind the offer of a solution to a different 'problem' that isn't a problem at all. There are many, for example, who argue along the following lines:

<u>The OD-OK argument</u>

P1: Problematic overdetermination requires that the causes be independent.
P2: Mental causes aren't independent of physical causes.
C: Mental causes and physical causes don't problematically overdetermine.

Different commentators offer different defences of P1 and P2, in particular by giving different accounts of the precise way in which mental causes aren't independent of physical causes. But at the heart of every version of the OD-OK argument is an appeal to the necessary connection between mental and physical causes– *without the holding of the necessary connection being explained*. To put the OD-OK move in its most basic form, in answer to the question 'why is there firing squad type overdetermination in every case of mental to physical causation?' the answer is supposed to be: 'because there is a necessary connection between the two causes, that is why they are always "overdeterminers". But this is not worrying: yes, *in a sense*, there is "overdetermination" but it is *not* troubling firing squad type overdetermination [i.e. it is not flukey]–there is a *necessary connection* between the causes, so they are "overdetermining" causes in a way that should not worry us.'

There is a long list of people who have put forward something like the OD-OK argument as at least a significant part of their response to what they perceive as the problem of overdetermination faced by NRP. For example: Goldman (1969: 481); Block (1990: 158-159); Garrett (1998: 371-372); Crisp and Warfield (2001: 314); Loewer (2002: 657; 2007: 257); Sider (2003: 4); Bennett (2003: 473-475; 2008b: 288); Kallestrup (2006: 471-473); List and Menzies (2009: 477n6); Kroedel (2008: 133-134; 2013: 12); Carey (2011: 261). Call these the 'OD-OKists'. They all appeal to the necessitation of the mental by the physical as a key part of their argument that physical effects are not problematically overdetermined by mental and physical causes. But in doing so, they're all simply addressing a *non*-issue, by, ironically, appealing to the very thing they *should* be addressing: an unexplained necessitation relation.

Indeed, not only is it a non-issue, it's not entirely clear that it's always expressed in way that's even coherent. The standard way of characterising the problem of overdetermination that NRP is supposed to face is to say that there's firing squad type overdetermination *systematically*, i.e. in every case of mental to physical causation there's flukey-overdetermination. But simply bolting on a systematicity condition like this creates a bit of a mess: saying that flukey-overdetermination holds in every case (of relevance to NRP) is just to say that there is a necessary connection holding between causes that don't stand in that necessary connection. This might seem too obvious an error for people to fall into, but this is what's being *said* when people talk about 'firing squad type' overdetermination that is *systematic* or that *happens in every case*–if they're intending to frame a problem for NRP. Karen Bennett, for example, counsels that NRPists should not respond to the overdetermination problem by:

> claiming that *systematic overdetermination of the everyday firing squad sort* is perfectly fine (Bennett 2008a: 3; my emphasis. The literature is littered with examples of similar comments, but for some prominent recent examples see Kim 2011: 215-216; Carey 2011: 253-254; Zhong 2011: 132, 132n4; Pereboom 2011: 129; Gibb 2013: 3; Kroedel 2013: 11).

This is what I call 'incoherent-overdetermination'. Charity demands, of course, that we should read the assertions of others in as coherent a manner as possible. So in order to see how charitable we must be here, let's consider the supposition in more detail.

Incoherent-overdetermination of a physical effect by a mental and a physical cause occurs iff those causes are:

1) distinct from each other;
2) each sufficient for the effect;
3) unmediated with respect to the effect;
4) standing in a necessitation relation to each other, and not standing in that necessitation relation to each other.

The first thing to note is that this is obviously not a problem for NRPists who can simply point out (if they've got nothing better to do) that their theory is obviously inconsistent with this. The problem of overdetermination purportedly faced by NRP is a problem faced by a physicalist position. Physicalism is a view which makes a *modal* claim about the mental and the physical, so when assessing it, we're interested in the set of worlds covered by that claim. So, to talk of *every* case of mental-physical causation is to talk about every case in that set of worlds. Therefore, to entertain the possibility that 'every case of mental to physical causation is a case of causal overdetermination' is to entertain a *modal* state of affairs, one in which every mental cause *must* be 'partnered by' a physical cause (within some range of worlds). But to entertain *that* is, of course, simply to entertain a necessary connection holding between the 'partner' causes. This is why it is incoherent to talk of 'firing squad type' overdetermination that *happens in every case*, or *systematically*.

Here's another way to think about it. The kind of case we're supposed to be considering is as follows: MC (a specific mental cause) stands in a *de re* dependence relation to PC (a specific physical effect), so, necessarily, if PC then MC. In this sense, PC must always be 'partnered by' MC–this is just what NRP entails. True, for any mental cause, MC_n, there will be any number of physical causes upon which it could depend, according to NRP; but, there must be such a physical cause and for any such physical cause, that physical cause will *de re* necessitate the mental cause in question. So, if we're working within an NRP framework, one cannot coherently suppose of any relevant MC-PC case that there is no necessary connection between MC and PC.

Surprisingly, no-one seems to recognise that incoherent-overdetermination is incoherent. Nowhere does anyone say ‘of course, firing squad type overdetermination that is systematic is incoherent’.[39] That slip might not be worth worrying too much about, but not only do people apparently make the supposition of incoherent-overdetermination as if it were coherent, they also appear to say that it is *the* problem of overdetermination faced by NRP. [40] But NRP is obviously inconsistent with incoherent-overdetermination, and so NRPists can bat away this ‘problem’ without batting an eyelid.

On behalf of those who seemingly make this error in trying to articulate Malcolm’s intuition, I think there are two likely responses.

First response: say that the overdetermination problem arises when mental-physical necessitation holds only *incompletely*. A *incompletely* necessitates B iff A necessitates B and there are worlds in which A does not necessitate B. The most obvious example of incomplete necessitation for our purposes might be nomic dependence. One could hold, for example, that the mental merely nomically depends on the physical *and* argue that nomically impossible but metaphysically possible worlds have some relevance for assessing the causal status of mental events which nomically depend on physical events. In that case, the purported problem would not be incoherent-overdetermination, but instead ‘incomplete-overdetermination’.

[39] The only possible exception that I can locate in the literature is when Kim states: ‘it would be incoherent to suppose there is another causal chain from M to P* that is independent of the causal process connecting P with P*’ (2005: 47-48). Here ‘P*’ is the physical effect and ‘M’ and ‘P’ are the supposedly competing mental and physical causes. Presumably, Kim thinks this would be incoherent in virtue of the dependence relation between M and P. Importantly, though, Kim uses this observation to motivate the claim that any M to P causal relation map on to the relevant P to P* causal relation, and *not* the claim that incoherent-overdetermination needn’t bother NRPists because their view is obviously inconsistent with it. Moreover, given what Kim says elsewhere about how he conceives of the problem of overdetermination faced by NRP, it would seem perhaps too charitable to grant Kim recognition of the role of incoherent-overdetermination as one of the unrecognised ambiguities of the various efforts to articulate Malcolm’s intuition. See below, especially Section 6, for more on this point.

[40] Although some do at least acknowledge that the problem is not supposed to be ‘systematic firing squad type’ overdetermination. Compare, for example, Gibb’s introductory chapter to a volume she co-edits (2013: 3) where she makes the error, with Árnadóttir and Crane (2013: 260) who, writing in the same volume that Gibb edits, explicitly try to grapple with a version of the problem which is not incoherent-overdetermination – namely incomplete or complete overdetermination (2013: 263), which I discuss below.

Incomplete-overdetermination of a physical effect by a mental and a physical cause occurs iff those causes are:

1) distinct from each other;
2) each sufficient for the effect;
3) unmediated with respect to the effect;
4) standing in an *incomplete* necessitation relation to each other.

Kim, for example, seems to have a variety of incomplete overdetermination in mind with his articulation of the overdetermination problem (2005: 49). If the necessitation holds incompletely (e.g. nomically), then it could at least be coherently stated that there are worlds in which it fails and so one cause occurs without the other; and given that there are those worlds, perhaps we have something *a bit like* flukey-overdetermination in the actual world because the overdetermining causes are independent in some modal sense (e.g. metaphysically). And if that were the case, then perhaps NRP would face a genuine problem articulated in the traditional way: 'if the mental "cause" had not occurred, then the physical effect would have occurred anyway–because it has a physical cause; thus the causal status of the mental "cause" is undermined–it makes no difference to the effect.'

However, there are two issues with this articulation. The first is that, in the absence of an argument that NRPists *must* hold that the physical incompletely necessitates the mental, this articulation targets a *version* of NRP, rather than NRP *full stop*. This immediately makes the super-overdetermination articulation preferable. Nevertheless, defenders of the incomplete-overdetermination articulation might still think they have a genuine problem, even if its scope is restricted; or they might think that they have a good argument that NRPists must hold that the physical incompletely necessitates the mental.

The second issue is that incomplete-overdetermination is not in fact a genuine problem. Suppose a theory entails that there's incomplete-overdetermination of physical effects by mental and physical causes. Next suppose God comes down from heaven and explains that He set up the world this way because He thought it was aesthetically pleasing. Do proponents of the theory have a problem? No. Why? *Because we have an explanation of* *why there is incomplete-overdetermination*.[41] Thus, the real issue is whether or not an explanation can be provided of *why* there is incomplete-overdetermination, not whether there *is* incomplete-overdetermination.

Here's another way to see the point. Consider flukey-overdetermination again: why are instances of it relatively mundane? Because we can *explain* why they occur: add in enough causal chains to a universe and it becomes increasingly probable that flukey-overdetermination will occur. It is not much of an objection to say at this point: 'ah, but it is possible for the flukey "causes" to come apart, so we need to downgrade the causal status of at least one of these "causes"'. Now consider incomplete-overdetermination, and grant for the purpose of argument that the necessitation is explained (either using my God example or a candidate naturalistic explanation of your choice). If that were the case, then it would not be much of an objection to say at this point: 'ah, but it is possible for the "causes" to come apart, so we need to downgrade the causal status of at least one of these "causes"'. Hence the real problem is not whether the causes can 'come apart', but whether we can explain the necessitation relation they stand in.

At this point, one might expect to hear the battle cry of all true metaphysicians everywhere: 'give me parsimony, or give me death!' I must confess that I too am moved by this call to arms. However, comrades, we must also be moved by the accompanying apothegm: 'not beyond necessity!' The question then is: how do we know how many causes are 'beyond necessity'? This is where the requirement for the explanation of mental-physical necessitation has particular purchase. We don't take Occam's razor to flukey-overdeterminers because we can explain those cases, so we should apply the same rule with incomplete-overdeterminers. True, we may yet find grounds for reaching for the razor, such as too many repeated attempts at explanation that fail, or a sound *a priori* argument that explanation is impossible. In such circumstances I'll join in the battle cry in defence of physicalism. And I offer no arguments here as to whether this outcome is likely or not; all I am maintaining is that *the NRP-overdetermination debate rests on whether or not an explanation can be given of the holding of the necessitation relation that the purportedly competing causes are said to stand in.*

[41] I don't think it matters, but if you want an explanation that is less obviously inconsistent with physicalism, just make it a physical alien 'god' from a super-advanced civilisation elsewhere in the weird and wonderful multiverse, who created this universe.

This brings us to the second response on behalf of those who talk about ‘systematic firing squad type overdetermination’. That response proceeds as follows: ‘there is no incoherence in talking about systematic firing squad type overdetermination because what people *really* mean (or *should* mean) is: the overdetermination problem faced by NRP is like firing squad overdetermination *except* that there is a *complete* necessary connection between the causes’. A *completely* necessitates B iff A necessitates B and there are *no* relevant worlds in which A does not necessitate B.[42] The most obvious example of complete necessitation for our purposes would be metaphysical dependence. If B metaphysically depends on A, then there are no worlds (let alone relevant ones) in which B doesn’t depend on A. However, one might argue that nomic dependence could also be a variety of complete necessitation, because one could defend the claim that all nomically impossible worlds are irrelevant.

Thus, nomic necessitation might be both complete and incomplete, in my sense. Because one could accept that there are metaphysically possible but nomically impossible worlds in which A doesn’t nomically necessitate B, and so accept that the necessitation is incomplete, but deny that those nomically impossible worlds are relevant, and so also accept that the relation is complete. Alternatively, one could argue that nomically possible worlds are all the worlds, and thus that nomic necessitation is complete and *not* incomplete. I offer these remarks not to recommend any particular approach. Rather, I intend only to illustrate how my notions of complete and incomplete necessitation come apart from the notions of nomic and metaphysical necessitation. I think the *complete* and *incomplete* categories are more useful because they capture more successfully the key difference between the traditional attempts at articulating Malcolm’s intuition: whether there are *relevant* worlds in which the necessitation fails. With that in mind, we can distinguish the following kinds of necessitation relations that are of interest:

Incomplete: A necessitates B, and there are worlds in which A does not necessitate B.

Relevantly incomplete: A necessitates B, there are worlds in which A does not necessitate B, and those worlds are relevant.

[42] By ‘relevant’ I just mean ‘relevant to the dialectic’. For example, one could argue that mental events are incompletely necessitated by physical events but that the worlds in which the necessitation doesn’t hold are worlds that matter to the debate. Or one could deny this. This is all I am referring to with my use of ‘relevant’.

Irrelevantly incomplete: A necessitates B, there are worlds in which A does not necessitate B, and those worlds are irrelevant.

Complete: A necessitates B, and there are no relevant worlds in which A does not necessitate B.

What, then, of complete-overdetermination? This sort of overdetermination of a physical effect by a mental and a physical cause occurs iff those causes are:

1) distinct from each other;
2) each sufficient for the effect;
3) unmediated with respect to the effect;
4) standing in a complete necessitation relation to each other (e.g. metaphysical dependence).

But notice how this purported problem is beginning to look suspiciously like the 'solution' proposed by the OD-OKists: P1 and P2 of the OD-OK argument entail that complete-overdetermination isn't a problem. But let's not be detained by that bump in the road. Because even if one can coherently add a systematicity condition to talk of flukey-overdetermination, and do this in a way that doesn't problematically pre-suppose one of the main solutions to the problem one is attempting to set-up, then, *at best*, we've simply been distracted by a red-herring. Because we can run the God argument again: if He stepped through the clouds and informed us that the world is thus and so in virtue of His penchant for the pleasing symmetry of complete-overdetermination, then there would be no problem–*because we would have an explanation of the holding of the complete necessitation*. Thus, again, the issue is whether or not an explanation can be provided of the complete necessitation, not whether there is complete necessitation. The real issue is, in other words, whether or not NRP can solve the *super*-overdetermination problem.

So, despite what so many say, or *seem* to say, flukey, incoherent, incomplete and complete overdetermination are not *genuine* problems *for NRP*, and so fail to meet the first articulation condition. Thus, we cannot say that any of them are *the* problem of overdetermination faced by NRP. Moreover, we struggle even to make sense of these attempts to articulate Malcolm's intuition, and in escaping that struggle end up back at super-overdetermination. This provides strong grounds for taking the super-overdetermination problem to be *the* problem of overdetermination faced by NRP.

This point is further strengthened when we consider in detail the best-known articulation of Malcolm's intuition, as put forward by Jaegwon Kim in his famous exclusion and supervenience arguments.

2.6 Jaegwon Kim's exclusion and supervenience arguments

Of all the contributors to the overdetermination debate Jaegwon Kim is by far the most prominent. His articulation of Malcolm's intuition has become the default, and in a series of contributions spanning decades, Kim brought the issue of overdetermination to centre stage and kept it there (see Kim 1979; 1984a; 1984b; 1988; 1989a; 1989b; 1990b; 1992a; 1992b; 1992c; 1993a; 1993b; 1995; 1996; 1997; 1998; 1999a; 2000; 2001; 2002; 2003; 2005; 2006; 2007a; 2007b; 2009; 2010; 2011). It will be worthwhile, therefore, to examine Kim's version of the problem of overdetermination, and put forward my criticisms of it.

In Kim's most recent attempts to articulate the problem, he presents two arguments, the supervenience argument and the exclusion argument (2006; 2009; 2011), that were previously run together under the title 'supervenience argument' (1998; 2002; 2003; 2005).[43] Let us focus on the most recent, two-argument version, which I think encounters a number of difficulties; namely:

[43] The one-part supervenience argument, in turn, was previously referred to with the terms: 'explanatory/causal exclusion' (1996; 1989a), or 'exclusion' (1993a), or 'explanatory exclusion' (1988, 1989b).

(i) in failing to recognise the real role dependence plays in the overdetermination problem, and instead employing the relation in ineffective ways.

(ii) in resting largely on the exclusion principle which is not 'virtually analytic' as claimed;

(iii)in failing to meet AC1;

(iv)in falling back on a productive account of causation which does not support the exclusion principle as intended; and at worst begs the question against NRP, or at best significantly reduces the force of the argument.

We will consider these criticisms shortly, but first let's get on the table some of the details of Kim's version of the overdetermination problem, which is best reconstructed with the following principles:

Closure: Every physical effect has a sufficient, unmediated physical cause.[44]

Exclusion: No unmediated overdetermination can occur unless it's flukey.

Distinctness: Mental events ≠ physical events.[45]

Supervenience: Every mental event depends on a physical event.

The exclusion argument employs the first three of the above principles, the supervenience argument the fourth. It will help us get to grips with the exclusion argument to do something Kim does not–assume the following 'preamble'.

Preamble

Grant NRP and thus that there will be instances of mental to physical causation in which a mental cause, MC, depends on a physical cause, PC, and both MC and PC are sufficient for, and unmediated with respect to, a physical effect, PE.

If you want to know why this preamble should be accepted, consider the following example. Alex's finger moves (a physical effect). When asked 'why did your finger move?' Alex says 'I felt a tingly sensation in it'. Thus, according to this psychological

[44] For Kim, 'physical' means 'microphysical' – where 'microphysical' means 'entity at the most fundamental level of reality or law that governs such' (2003: 173). This contrasts with my use of 'physical' to mean 'concrete non-mental entity or law that governs such'. For the current dialectic, when I attribute to Kim the use of 'physical' it should be read in my sense; this does not impact significantly on the debate.

[45] Note that in the metaphysic Kim works with, events are property instances – ordered triples of an object, property, and time (o, F, t). This is why, in Kim's model, mental events are distinct from physical events if even the most minimal form of NRP (standardly conceived) is granted in which only mental properties are distinct from physical properties.

explanation, a psychological event caused the movement. We then ask a neurophysiologist, who's conveniently standing nearby, 'why did Alex's finger move?' and she gives a physical explanation involving brain activity, electrical signals and muscle contraction. Thus, according to this neurophysiological explanation, a physical event caused the movement. NRP asserts that mental events are not identical to physical events, so Alex's sensation is not identical to any neurophysiological event; but NRPists also assert that all mental events depend on physical events, and in particular cases like Alex's finger assert that the sensation depends on the neurophysiological event. Thus, NRP entails that there is a physical cause, PC, and a mental cause MC, such that MC depends on PC and both MC and PC are sufficient and unmediated causes of the physical effect, PE.

Kim tries to work much of what's going on in the preamble into the early stages of the exclusion argument, but the results invite criticism (i), a consideration of which invites, in turn, criticisms (ii) through (iv). To see why, consider the exclusion argument proper, which takes the form of a reductio:

<u>Exclusion argument</u>

P1: A mental event, MC, causes a physical event, PE (*assumption; or from preamble*).
P2: PE has a physical cause, PC (*from closure and P1; or preamble*).
P3: MC ≠ PC (*from distinctness; or preamble*).
P4: MC and PC causing PE isn't a case of flukey-overdetermination (*from preamble*).
P5: PE can't be caused by MC *and* PC (*from exclusion and P4*).
P6: PE must be caused by PC (*from closure or P2*).
C: MC doesn't cause PE (*from P5 and P6*).

The argument can be run for every purported instance of mental-physical causation, so if one wishes to accept that mental causes have physical effects then one must deny one of the premises.

Criticism (i) amounts to pointing out that there is a problem in Kim's setting up of the overdetermination problem in the ways he employs dependence. My over-arching argument has been that everyone has been setting up the overdetermination problem in the wrong way by failing to properly recognise the role that the particular necessitation relation plays in the problem. It seems worthwhile, therefore, to look in some detail at how the most famous attempt to set up the problem repeatedly fails to employ dependence correctly in trying to construct a genuine problem for NRP.

Begin by noticing that if we read the exclusion argument *without* the preamble (as Kim would have it), it is not obvious why, by Kim's lights, we should accept that P4 is true. The most natural justification for P4 would be to say that MC *depends* on PC, but in Kim's most recent version of the argument he conspicuously fails to say this, and simply asks the reader to grant P4 (2011: 216; 2009: 39; 2006: 197).

I cannot help but find this failure to employ dependence *at all* in the exclusion argument pretty mysterious. It must be that Kim just thinks he doesn't need it. And, indeed, he does offer other reasoning when introducing the argument that isn't explicitly aimed at supporting P4, but which could support it. Namely, that if we take MC and PC to be flukey-overdeterminers of PE, then this undermines the status of MC as a 'full and genuine' cause of PE, because PE would have occurred anyway even if MC hadn't occurred. As he states:

> '[a case in which] a physical cause, even if the mental cause had not occurred, would have brought about the physical effect […] appears to weaken the status of the mental event as a cause of the physical effect. To vindicate *m* [the mental cause] as a *full and genuine cause* of *p* [the physical effect], we should be able to show that *m* can bring about *p* on its own, without there being a synchronous physical event that also serves as a sufficient cause of *p*' (2011: 215; emphasis added).[46]

I think there is a plausible reading of what Kim could mean here by '*full and genuine cause*', which would justify P4. That reading: 'genuine' means 'real', and 'full' means 'at least necessary'. Where 'necessary' would mean that *given enough of the context being fixed*, the effect could not have occurred without the cause. Hence, we should accept P4 because causes should be necessary for their effects, and in instances of flukey-overdetermination each 'cause' is not necessary for the effect, because if one cause is taken away the effect still occurs. So if we want to say that MC and PC are causes, then we can't say that there's flukey-overdetermination here.

[46] In his 2006, Kim has the same passage except 'full and genuine' is simply 'genuine'.

This reading would justify P4: MC and PC causing PE couldn't be a case of flukey-overdetermination because if it were, then MC couldn't be a full and genuine cause of PE, and by stipulation it is. This reading is also supported by the fact that Kim assumes causes must be necessary for their effects in P3 of the supervenience argument, as we will see below.[47] Unfortunately, however, we can't just leave matters there. Because, if causes must be necessary for their effects then in cases of flukey-overdetermination neither purported cause qualifies as a genuine cause. However, Kim does state that there *can* be genuine causes in instances of flukey-overdetermination, when he allows for:

> genuine overdetermination [as] illustrated by the "firing squad" example (ibid: 216).

So Kim must think that flukey-overdetermination is possible for firing squads but *not* for mental and physical causes–in the cases of interest, anyway.[48] I agree, of course, that flukey-overdetermination doesn't occur in the type of mental-physical case of interest, *because such cases involve a dependence relation between the causes*. But, puzzlingly, Kim doesn't appeal to mental-physical dependence to justify P4 when framing the latest version of the exclusion argument, leaving us only with the thought that there can't be flukey-overdetermination in the mental-physical case because causes must be necessary for their effects. This makes Kim's acceptance of flukey-overdetermination in the firing-squad case surprising–because neither of the shootings are necessary for the death.[49]

This brings us back to the best reason to accept P4: that MC depends on PC; in other words we're just considering the kind of case which isn't a case of flukey-overdetermination. Kim did give this reason previously (2005: 48); but, for reasons that are not entirely obvious, Kim doesn't say this in his more recent formulations of the argument. I suspect that this might be because Kim has previously put forward another version of the exclusion argument which is consistent with there being no dependence between MC and PC; Kim calls this version 'completion 2' (2003: 158-159: 2005: 44).[50] Perhaps, therefore, in his

[47] Kim also endorses a '*sine qua non*' condition on causes elsewhere (2007a: 235).

[48] Here, we are not concerning ourselves with mundane cases of mental-physical flukey-overdetermination such as: you intend to lift your arm (mental cause) and a rope simultaneously lifts your arm (physical cause) and your arm goes up (physical effect).

[49] Moreover, given Kim's preference for the incomplete-overdetermination articulation, and acceptance of flukey-overdetermination cases (presumably because they are so easily explainable) he faces a particularly pressing question: if flukey-overdetermination is acceptable because it's explainable, then why rule out incomplete-overdetermination in the absence of an argument that it is not explainable?

[50] Charity, of course, might require that we simply grant Kim an appeal to mental-physical dependence to justify P4. But that the move seems so obvious makes it especially puzzling why Kim doesn't just come straight out with such an appeal when framing the exclusion argument in its most recent form.

most recent version of the exclusion argument that we are considering here, where Kim also fails to say that P4 is justified by MC depending on PC, Kim has completion 2 in mind. The problem with completion 2, though, is that its relevance for NRP is unclear. The cases of overdetermination that we're interested in just are those NRP supposedly entails: non-flukey cases where the mental cause depends on the physical cause.

So, in his three most recent statements of the exclusion argument (2006; 2009; 2011), Kim refuses to appeal to dependence at all, even when it would seem natural to do so to justify P4. In his earlier 'completion 2' variations of the argument, Kim also doesn't appeal to dependence to justify P4 (2003: 158-159; 2005: 44). But allowing MC not to depend on PC would make the case irrelevant to the NRP-overdetermination dialectic. This makes it difficult to interpret what is motivating Kim when he formulates the exclusion argument without appealing to dependence.

In an even earlier version of the exclusion argument, Kim employed dependence in trying to set up the argument in a further, rather convoluted way. Kim proceeded as follows: MC causes PE (by assumption), and MC must have a subvenience base, PC (from supervenience). At this stage, Kim hasn't established that PC is also a cause of PE, and so argues for this as follows: PC must also cause PE because A causes B if it's the case that: either A is nomically sufficient for B, or B counterfactually depends on A; and, PC *is* nomically sufficient for PE, and PE *does* counterfactually depend on PC. In this way: PC is nomically sufficient for MC (from supervenience), and MC is also nomically sufficient for PE (by assumption), so by transitivity, PC causes PE. Similarly, PE is counterfactually dependent on MC (by assumption), and MC is also counterfactually dependent on PC (from supervenience), so–again by transitivity–PC causes PE (see Kim 1998: 43; similar reasoning is gestured towards in his 2005: 41).

One immediate problem with *this* way of employing dependence is that it involves an invalid inference. First, Kim says that PC causes PE *via* MC, before then concluding that PC *directly* causes PE. But we haven't been given a reason to think that PC does directly cause PE. And if we simply run with the set-up whereby MC directly causes PE, and PC causes PE via MC, then we have an implausibly complicated case whose relevance to the NRP-overdetermination dialectic is not clear. Consider: we would have, first, a case of instantaneous causation (which some people won't like): PC causing MC. Second, an instance of an event, PC, being both a cause of, and the subvener for, another event, MC.

And third, a peculiar instance of overdetermination in which the two causes are synchronous, PC and MC, but only one, MC, is unmediated. In my view, if we can avoid such complications, we should; not least because these complications don't even deliver a case relevant to the NRP-overdetermination dialectic because it's just not the kind of case that NRP entails.[51]

The good news is that we can avoid all this palaver by just granting the preamble, i.e. by stating that the case we're looking at just is the case that NRP entails: a mental and a physical cause that stand in a dependence relation and that overdetermine a physical effect.

Note, also, that if we grant the preamble, it makes most of the exclusion argument proper superfluous–indeed all that's left is P5 and P6. The exclusion premise, P5, is where the action's at, so let's focus on my rejection of it, which constitutes criticism (ii): that Kim errs by resting the exclusion argument largely on the exclusion principle.

First let's get clear on what the principle is. In Kim's own words, the exclusion principle states:

> No event has two or more distinct sufficient causes, all occurring at the same time, unless it is a genuine case of overdetermination (ibid: 216).

I think this is best read as ruling out unmediated overdetermination in all cases except flukey ones. Kim thinks the exclusion principle is next to impossible to deny. He describes it as 'obviously, almost trivially, true' (ibid), and 'virtually an analytic truth without much content' (2005: 51; 2003: 163). The reason for this confidence is not entirely clear. The exclusion principle would be analytic under one reading, say: 'no synchronous overdetermination unless there's synchronous overdetermination'. There's certainly not much content in *that* principle. But a tautology is not much use in ruling out the overdetermination cases NRP entails. Perhaps what's driving Kim's thinking here is that exclusion is *close* to this because it states: 'no synchronous overdetermination unless it's this one specific kind of overdetermination: flukey.' I can *perhaps* see how, when considered in relation to the actually analytic, trivial version, this might *seem* '*almost*

[51] Indeed, puzzlingly, Kim appears to argue against himself when he contends that PC *couldn't* cause PE via MC for reasons similar to those I give (2003: 157; 2005: 42).

trivially true' or '*virtually* analytic'–it's similar to the trivial, analytic version, but yet is not completely vacuous.

But, as we have seen, a theory entailing incomplete or complete overdetermination only faces a genuine problem in the absence of an explanation of the necessitation relation the causes stand in to one another. Thus criticism (ii) leads us to (iii): in that Kim erred in resting the exclusion argument on the exclusion principle because in doing so Kim has failed to meet AC1. In order to meet AC1, NRPists would at least have to face some pressure to accept the exclusion principle as stated, but they don't. Indeed, this can be seen if we summarise the exclusion argument as follows:

<u>Exclusion argument summary</u>
P1: Preamble.
P2: The exclusion principle says the preamble is impossible.
C: The end–for NRP.

Unsurprisingly, therefore, by far the most popular move NRPists make in responding to Kim is to deny the exclusion principle. This often involves an OD-Okist style move, or lengthy ruminations over the nature of causation. But, as I have shown, these have been efforts towards a solution to a non-problem; here, in Kim's exclusion argument, we see one of the main sources for the confusion. The force of the exclusion argument rests largely on the shoulders of the exclusion principle, which rules out incomplete or complete overdetermination without good reason; hence, Kim's exclusion problem is not a genuine problem, and so fails the first articulation condition.

That we can't simply rule out the possibility of overdetermination without argument was noted at the dawn of this debate. In the first response to Malcolm's original paper, Alvin Goldman pointed out that it is:

> at least conceivable that all human movements have two sets of sufficient conditions, one involving neurophysiological states and one involving desires or intentions. An argument is needed to prove this is not, or could not be, the case (1969: 469; for a similar point, see Martin 1971: 86).[52]

[52] Indeed, Malcolm himself also considers this possibility, but ultimately rejects it (1968: 52).

It is interesting to note, therefore, that in Kim's articulation of Malcolm's intuition, this issue remains very much alive: that there cannot be such overdetermination is not argued for, instead it is treated as 'obvious', or 'virtually analytic'.

Or, at least, this was the case until relatively recently. In an effort to justify the exclusion principle, since the turn of the century Kim has fallen back on a productive account of causation (henceforth: 'productionism'). This brings us to criticism (iv): that in relying on productionism, Kim fails to support the exclusion principle, and, at worst, begs the question against NRP, or, at best, significantly reduces the force of the exclusion argument.

To see why, it will be helpful first to get clear on how Kim's views of causation have developed alongside the development of the NRP-overdetermination dialectic. In what follows I will discuss some simple versions of causal theories that face a variety of well-known objections. Those debates do not concern us here; I discuss these simple versions to illuminate the current dialectic only.

For much of his career, and during a period when the approach enjoyed some popularity, Kim held a nomic subsumption account of causation (1973a; 1973b; 1974; 1977; 1984a; 1984b; 1988). The account, which is a variant of the Humean regularity approach, holds that causal relations between events should be analysed as instances of lawful regularities:

C is the cause of E iff there is a law which subsumes them, C≠E, C and E occur, and C is spatiotemporally proximate to E.[53]

This was the view of causation Kim had in mind when first formulating the problem of overdetermination purportedly faced by NRP. So, for instance, when Kim (in the twentieth century) talks of *causal sufficiency* he means something like:

C is causally sufficient for E iff the (correctly predicating) sentences 'C occurred' and 'L holds' entails 'E occurred'.

[53] Recall that a law, L, subsumes events *C* and *E* iff there are predicates F and G such that *C* satisfies F and *E* satisfies G, and L implies $\forall x(Fx \rightarrow Gx)$; and '$\forall x(Fx \rightarrow Gx)$' is not a logically necessary truth. For example, the law: *all pure water boils at 100 degrees centigrade in Earth's atmosphere*, subsumes the events: *this pure water was heated to 100 degrees centigrade at t*, and: *this pure water boiled at t*. Thus, nomic subsumption is a species of nomic implication between sentences containing correct event predications.

Kim still thinks that some version of the nomic subsumption account might be true at the most fundamental level of reality–once we get down there, all we can find are lawful regularities (2007a: 232). But at higher levels, particularly the psychological level, Kim now holds productionism. The first explicit acceptance of it by him in print occurred this century (2002: 675), although there are earlier signs that he was attracted to the view (1997: 286; 1998: 45). Initially, Kim did not spell out at any great length what he meant by 'productive causation' (2002: 675-677; 2005: 18, 47n12), although he has given a fuller account of his views more recently (2007a: 227-239; 2009: 44-45).[54]

So what is productionism? To give an illuminating answer to this question it will be helpful to compare productionism with its *bête noire*: the counterfactual approach. These two theories of causation are thought by many to reflect a significant divide in our commonsense causal concept (particularly following Hall 2004). Also, Kim's adoption of productionism is in significant part driven by his rejection of the counterfactual account, which he has been critical of since Lewis first proposed it (Kim 1973b), and has remained consistent in rejecting throughout his career (e.g. Kim 1997; 1998; 2002; 2005; 2007a; 2009).

According to a simple version of counterfactual account:

C is the cause of E iff E counterfactually depends on C, C≠E, and C and E occur.

Productionism comes in different varieties, but the most discussed, and the one Kim prefers, is the conserved quantity view, according to which:

C is the cause of E iff a quantity of C is transferred to E, C≠E, C and E occur, and C is spatiotemporally proximate to E.

[54] Some prominent commentators have missed this more recent and more expansive explanation of Kim's. Here are two examples. First: Christian List and Peter Menzies have rested their criticisms of Kim in part on charging him with causal theory obscurantism – a charge which is at least somewhat unjustified given Kim's 2007a article, which they do not cite. List and Menzies suggest that: 'unless a better explication can be given of causation as production, this notion can hardly play a significant role in the debate about mental causation' (2009: 489). Second: Steinvör Thöll Árnadóttir and Tim Crane suggest that there is a 'lack of argument' to support Kim's exclusion principle (2013: 259-262). However, they fail even to mention Kim's appeal to productionism, despite, peculiarly, citing (2013: 253, 263, 265) Barry Loewer's 2007 chapter, in which Loewer criticises Kim's productionism in response to an immediately preceding chapter in which Kim defends his productionism, in a volume entitled *Contemporary Debates in Philosophy of Mind* (see Kim 2007a & Loewer 2007).

The quantities most often entertained as counting as *causal* quantities include *energy-mass*, *charge*, or *momentum* (see Fair 1979; Salmon 1998; Dowe 2000). To illustrate the production/counterfactual contrast consider a case where a white billiard ball fails to hit a red ball which is already in motion heading towards the left corner pocket, and the red ball then goes into that pocket. But if the white ball had hit the red ball, the latter would have gone into the right pocket. In such a case, we might say that white ball 'caused' the red to go into the left pocket rather than the right. The counterfactual approach would allow cases of 'omission' like this as an instance of causation, the conserved quantity view wouldn't as the white ball doesn't transfer any quantity to the red. The productive account restricts itself to cases such as when the white ball *hits* a red ball thereby *producing* motion in it–the quantity of momentum is 'passed' from one ball to the other (although one ball hitting another would also count as causation on the counterfactual view).

Kim opts for productionism because he believes that what's really at issue when considering mental causation is the question of whether or not mental causes *produce* physical effects as this is what genuine human agency requires. His argument proceeds as follows: nomic regularities can't be identified with causation because we want to know which nomic regularities are causal, and which aren't. But we can't know *this* without 'digging deeper' to find some other feature of the world which isn't also a 'mere' nomic regularity (unless we're at the fundamental level, in which case perhaps we'll simply have to make do with the regularities we find there, as there is no deeper place to dig to). Thus, this need to dig deeper tells us that nomic regularities can't be identified with causation. Also, counterfactual dependence can't be identified with causation because patterns of counterfactual dependence in turn are determined by nomic regularities–we assess the truth of counterfactual claims by appealing to the laws that hold in a world. Kim also worries that patterns of counterfactual dependence can be accepted by epiphenomenalists–to use one of his examples, consider the shadows cast by a moving car. A car shadow at t_1 does not cause a car shadow at t_2, although the counterfactual 'if there had not been a car shadow at t_1, then there would not have been a car shadow at t_2' may well be true. The general point would seem to be that patterns of counterfactual dependence might imperfectly 'track' *real* causal processes even though the counterfactual dependence relation itself isn't the causal relation. Thus, if we simply point to the patterns of counterfactual dependence that seem to hold in the case of mental causation, we can't safely conclude that we aren't simply pointing at shadows.

Kim thinks this leaves us with only one choice if we're concerned about human agency:

> An agent is someone who […] has the capacity to perform actions in the physical world: that is, to cause her limbs and other bodily parts […] to move in appropriate ways […] It seems to me that without productive causation, which respects the locality/contiguity condition, such causal processes are not possible. These causal processes all involve *real connectedness* between cause and effect, and the connection is constituted by phenomena such as energy flow […] an actual movement of some (conserved) physical quantity (2007a: 236).

I cannot pretend that I am persuaded by this line of thought. But others see more force in it. For example, Barry Loewer–a critic of Kim–believes that productionism would at least support the exclusion principle:

> Kim seems to think of causal production as an intrinsic relation between relatively local events. So, for example, a brain event *produces* bodily motion by transferring some kind of causal energy, or, as Hartry Field […] says, 'causal liquid', from the cause to the effect. *It seems obvious* that if the brain event produces the bodily motion, a distinct mental event has *nothing more to do* (2007: 253; emphasis added).

I will not resist this by engaging in the debate on how plausible productionism is in general, or the conserved quantity view in particular. Indeed, I think if one can avoid such a debate in one's articulation of Malcolm's intuition, then so much the better. But I can at least see how this might undermine the kind of OD-Okist response we considered above, although what is harder to see is how productionism is supposed to *support* the exclusion principle. A lot hangs on what quantities productionism allows to be transferred. If the kind of overdetermination supposedly entailed by NRP is one where there are two concrete causal processes of physical energy transfer going on, then NRP is obviously false. But that would be because the theory of causation would entail that there can only be physical quantities transferred, and mental causal processes, as such, don't transfer physical quantities, as such, so there would be no mental causation, as such. But that's hardly going to persuade NRPists, and rather than *supporting* the exclusion principle it would simply make it *superfluous*–the causal theory would be doing all the work.

Kim recognises this question-begging issue, but does little to address it–suggesting, instead, that it is NRPists who face the burden of explaining how their view is compatible with productionist mental causation (2007a: 238-239). I think this is, from a dialectical perspective alone, deeply unsatisfactory.

So, it is tempting to consider if a more liberal understanding of what a 'quantity' is might help Kim here. For example, if the quantities allowed by the causal theory include mental quantities (whatever they might be), then the theory might not be obviously inconsistent with NRP. However, this would then undermine the thought that productionism supports the exclusion principle, because it would then appear that NRPists face no more of a problem if their theory entails incomplete or complete overdetermination of the productive sort versus the counterfactual sort. *Perhaps* the former might *seem* harder to explain away than the latter, although the reasons why are not immediately obvious. Loewer says it's because it would entail that the mental event would have '*nothing left to do*'–but that phrase does little more than gesture again at the intuition the causal theory is supposed to be elucidating. There's a sense in which the mental cause has 'nothing left to do' if productionism is assumed, and there's a sense in which the mental cause has 'nothing left to do' if the counterfactual view is assumed. In the latter case, to be sure, OD-Okists have appealed to mental-physical necessitation (of one sort or another) to frustrate efforts to spell out the overdetermination worry by saying things like 'well, if the mental cause hadn't occurred, then the physical effect wouldn't have occurred, so in that sense the mental cause doesn't *do anything*.' OD-Okists respond by saying, 'in the relevant set of worlds, it's impossible for the two causes to come apart.' I suppose *that* response of the OD-Okists can be frustrated by this appeal to productive causation, because the worry about overdetermination would be expressed in non-counterfactual terms.

Nevertheless, whether we want to rule out productive overdetermination will *still* depend on whether an explanation can be provided of why the overdeterminers stand in a dependence relation. This is because if productionism isn't to be straightforwardly incompatible with a non-reductionist model, then it can't rule out physical-physical overdetermination, so it's not clear on what basis it could rule out mental-physical overdetermination.

Consider: on a non-reductionist model, we can have a situation where a biological event, BE, depends on a chemical event, CE, and there's some chemical effect, Ceff, such that BE and CE synchronously overdetermine Ceff. For example, a petri dish is filled with a solution such that, when the cell in it divides (BE), the oxygen level in the solution increases (Ceff). And the cell is of course composed of carbon, hydrogen, oxygen and nitrogen elements (CE). So, we can ask, was it BE or CE that caused the Ceff? Importantly, accepting the productive account of causation here doesn't add much to the mystery *assuming it's compatible with this set-up in the first place*. Let's frame the thought in terms of conserved quantities. We obviously can't rule out biological quantities, because this debate is supposed to be happening within a physicalist framework–everything is at least physical or depends on it. So if the account of causation accepts physical quantities, as it surely must, then there are biological and chemical quantities. So we need to allow that there can be biological and chemical quantities–both kinds of 'causal juice'. Now run the issue again with a counterfactual account of causation–what's changed? Does it suddenly look any more or less *weird*? And if you have an intuition either way, how exactly has that moved the debate forward? For what it's worth, my answers to these questions are, respectively, 'not much', 'my intuitions are not clear', and 'it has not'.[55]

[55] It might be suggested that with productive overdetermination, there must be an increase in quantity in the effect that would not otherwise be there in the absence of the 'extra' cause – i.e. there would an excess of causal juice. But this needn't be the case; for example, the causal juice of overdeterminer A could form a proper part of the causal juice of overdeterminer B. In that case, we would be none the wiser, as the amount of causal juice in the effect would show no excess. This is another demonstration of the crucial role played by the relationship between the overdetermining causes in apparently problematic cases of overdetermination.

Compare this situation now to the mental-physical case: MC and PC overdetermining PE. Given a conserved quantity account of causation, are we in a better position to rule out mental quantities than we were to rule out biological quantities? It's at least true that mental quantities aren't *physical*, but it's not true, if we're physicalists, that they're not what I call '*broadly physical*'–something is broadly physical iff it's physical or depends on the physical. The question then becomes: does the causal theory rule out broadly physical mental quantities which aren't physical? If it does, then it simply begs the question against NRP. If it doesn't, then what's the important difference between the mental-physical case and the biological-chemical case? It's true that in the former we have a physical cause (PC) and a broadly physical mental cause (MC) overdetermining a physical effect (PE), whereas in the latter it's two physical causes and a physical effect. But if *this* is what the 'obvious', 'virtually analytic' truth is supposed to be (that you can't have a physical cause and a broadly physical mental cause synchronously overdetermining a physical effect), then it is very far from obvious why anyone would think it *virtually analytic*.

Thus, under pressure to justify the exclusion principle, Kim has adopted a strong view of the causal relation in return for benefits of dubious value. Indeed, adopting productionism has weakened his articulation because, at best, it narrows its generality. Moreover, this has all been done in an attempt to shore up an articulation of Malcolm's intuition that fails the first articulation condition: even granting a productive account of causation, beyond the rather indefinite extra intuitive 'weirdness' of productive overdetermination, there is no *real* problem without the unexplained dependence between the two causes. The *problem* remains, in other words, ruling out *super*-overdetermination.

To close this section, let us come full circle and return to criticism (i): namely that Kim failed to recognise the real role dependence plays in the overdetermination problem, and instead employed the relation in ineffective ways. A further example of this can be found in the supervenience argument, where Kim gives dependence, under the label 'supervenience', yet another role to play in addition to those we have discussed above. Compare the part dependence plays in the super-overdetermination problem to its part in the supervenience argument, the latter of which endeavours to generalise the exclusion problem to all instances of mental causation:

Supervenience argument

P1: A mental event, MC, apparently causes another mental event, ME.

P2: ME depends on a physical event, PE (*from supervenience*).

P3: If P2, and MC doesn't cause PE, then MC doesn't cause ME.

C: MC can cause ME only if MC causes PE (*from P3*).

Thus, the exclusion argument concludes that, given NRP, mental causes can't have physical effects, and the supervenience argument concludes that supposed mental causes can have mental effects only if they cause the physical events on which the mental effects depend. In this way, the latter argument generalises the result of the former to cover *all* mental causes. Because, when looking at ME, Kim suggests, there are two answers to the question 'why did ME occur?' First, MC caused it. Second, PE necessitated it. But, he continues, PE necessitating ME undermines the status of MC as the cause of ME, because MC would not then be necessary for ME in the relevant 'context fixed' sense. Thus, if MC *is* to be so necessary, then it had better cause PE. Thus, to generalise, mental events can be causes only if they cause physical events.

So, in Kim's articulation, the problem isn't that there's an unexplained necessitation relation: dependence; it's that the dependence forces what he calls 'downward causation' from mental causes to physical effects in every case. Thus, although Kim sees dependence as an important part of the overdetermination problem, he views it as exacerbating the problem he believes he has identified with the exclusion argument. This contrasts quite significantly with the central role dependence plays in the super-overdetermination problem.

It is this failure to comprehend the fundamental role the unexplained dependence relation plays in the issue that has undermined the NRP-overdetermination dialectic. I will now show why reductive physicalism also faces a problem which in many ways echoes the problem faced by NRP.

2.7 Reductive physicalism and overdetermination

If you find yourself conversing with some philosophers of mind, and you say ‘it is not just non-reductive physicalism that faces an overdetermination problem, reductive physicalism faces a similar problem too’, then you will likely cause some surprise. Because to say that reductive physicalism faces an overdetermination problem anything like the problem faced by its non-reductive nemesis is to push against a deeply-entrenched consensus.

I wish to convince you that it is quite clear, upon reflection, that reductive physicalism faces, in some sense, an overdetermination problem that is similar to the super-overdetermination problem we have been dealing with in these pages. To do this I will answer three questions. First: what is reductive physicalism? Second: why might one think that it *doesn’t* face anything like NRP’s overdetermination problem? Third: why should one believe that it *does* face something like that problem?

The first two questions we can answer relatively quickly: as we have already seen, reductionism denies the distinctness tenet of NRP, and gives a different interpretation to the nothing above tenet–accomplishing both in one move: by holding that mental properties and events are identical to physical properties and events. And, on this reductionist model, it looks like the possibility for any sort of causal ‘competition’ is ruled out straightaway: there are no distinct mental causes which can ‘compete’ with physical causes–mental causes are *identical* to physical causes: something can’t causally compete with itself!

Consider the super-overdetermination articulation of Malcolm’s intuition that we have been working with. Super-overdetermination of a physical effect by a mental and a physical cause occurs iff the causes are:

1) distinct from each other;
2) each sufficient for the effect;
3) unmediated with respect to the effect;
4) and the particular necessitation relation they stand in to each other is metaphysically brute.

This clearly pre-supposes *two* causes–which reductive physicalism is obviously inconsistent with. And you need two causes in order to have a problem of overdetermination, right?

Well, maybe not. First, remember the original intuition: an appearance of competition between two kinds of causal *explanation*, 'mechanistic' versus 'purposive'. And here's a way of articulating that intuition which meets AC1 and AC2 and *presents a problem for reductive physicalism*; although, importantly, this is a problem of NRP too. I call the problem 'epistemic-overdetermination'.

Epistemic-overdetermination of a physical effect by a mental cause, A, and a physical cause, B, occurs iff:

1) A and B are distinct for all we know;
2) A and B are each sufficient for the effect;
3) A and B are each unmediated with respect to the effect;
4) and the particular necessitation relation A and B stand in to each other is epistemically brute.

The particular necessitation relation in question for reductionists is identity. Recall that an epistemically brute identity is a claim of the form 'A=B' where it does not appear that A=B and no true, informative explanation can be given for why we should believe that A=B. It is important to get clear on what this condition amounts to in the present context.

First, note that if everyone in the philosophy of mind agrees on anything, then they agree that there *appears* to be a distinction between the mental and physical. Indeed, this appearance, and our efforts to see beyond it to whatever metaphysical truth lies beneath has been the core concern in the philosophy of mind since there has been an area of enquiry deserving that label. And this is where Malcolm's worry begins: an *epistemic* difference between two kinds of causal explanation–mental and physical. In the super-overdetermination articulation we followed the standard practice of granting that those two explanations refer to distinct mental and physical causes. But we can articulate Malcolm's intuition without doing that. Indeed, Malcolm himself recognised this, ending his original article with the hope that:

> Perhaps the publication of the present paper will be justified if it provokes a truly convincing defence of the compatibility of the two forms of explanation (1968: 72).

And this, I think, expresses the most basic challenge posed by Malcolm's intuition: *demonstrating that the two types of explanation are compatible*. We have already seen in detail how difficult a challenge that is for *non*-reductive physicalism. But reductive physicalism also has to demonstrate the compatibility; and, crucially, it is not enough simply to say 'ah, well, reductionists deny distinctness, so they *can't* face an overdetermination problem'. Because simply asserting that a particular necessitation relation holds won't do, whether that necessitation relation is identity, or anything else. The holding of the identity relation, just like the holding of the dependence relation, *has to be explained*–and *that's* the *problem*.

We can even outline analogous explanatory criteria for explaining the holding of identity: that any relation R, explains the holding of mental-physical identity only if it is shown that:

(i) R entails identity;
(ii) R is internal;
(iii) R holds between mental and physical events.

Here, R will have to be identity itself because no relation is stronger than identity, so the first criterion is trivial.[56] Relatedly, reductionists don't face any Humean concerns about necessary connections between distinct entities. They have, by their own lights, no distinct entities to worry about. Non- reductionists do, and so the explanatory demands facing them come, perhaps, with the extra pressure of dealing with the Humean intuition that necessary connections between distinct entities are metaphysically suspicious.

Nevertheless, there is a broader similarity between the explanatory challenge faced by reductive and non-reductive physicalism. For non-reductionists the challenge is: 'you say the mental stands in R to the physical, but why believe that it does?' And for reductionists it is: 'you say the mental is identical to the physical, but why believe that it is?'

A natural response at this stage might be to press the point that we saw Block and Stalnaker express in Chapter 1, and say that asking why something is identical with itself is

[56] Excepting trivial counterexamples like the relation of *being identical to and being the same colour as*.

simply wrongheaded. But recall the distinction we drew: (i) whether it makes sense to ask why A=B; (ii) whether it makes sense to ask why we should believe that A=B. Thus, the explanatory challenge faced by reductionists is of the latter not the former kind–i.e. reductionists are *not* being expected to answer: 'why is A identical with A?' Instead, the challenge is answering 'there seems to be A, and a distinct B, you reductionists say A *is* B; but why should we believe that A is B when A and B do not appear to be identical?'

Here's another way to see the point. Alex looks at the Morning Star, then later that day he looks at the Evening Star. You tell Alex 'the Morning Star is identical to the Evening Star'. Alex replies 'but they don't *look* identical–one of them appears only in the morning, and the other only in the evening..?' You then give Alex an explanation of the solar system, the planets in it, their orbits, and thereby demonstrate that the planet Venus appears in the exact position that the Morning Star and the Evening Star appear in; thus, given that the Planet Venus is one object, and it is identical to the Morning Star and the Evening Star, then the Morning Star is identical to the Evening Star–they're both the planet Venus.

With regards to the epistemic-overdetermination problem, reductionists are in a similar position. Reductionists say 'the mental cause is identical to the physical cause'. Non-reductionist Alex replies 'but they don't *look* identical–it seems possible to have the mental cause in the absence of that specific physical cause'. The reductionists then need to explain to Alex why he should reject appearances and accept the identity.

Here's a question: which problem is harder? The super-overdetermination problem, or epistemic-overdetermination? Non-reductionists face both–but if they can solve the super-overdetermination problem, then they'll solve the epistemic-overdetermination problem, the solution to the latter will 'come for free' with the solution to the former. Thus, there is an important sense in which the epistemic-overdetermination problem is primarily a problem of reductionists. It might look like non-reductionists have the harder task with super-overdetermination, particularly if one shares the Humean intuition. But on the other hand, only reductionists come up against what might be called the 'Cartesian intuition'– that the mental and the physical are distinct. So, you pays your money, and you takes your choice. You can choose a non-reductionist model that automatically accommodates the Cartesian intuition of mental-physical distinctness, but requires a necessitation relation that worries the Humeans and the holding of which seems very difficult to explain. Or you can choose a reductionist model which certainly won't worry the Humeans, but that doesn't accommodate distinctness, and thus requires a persuasive anti-Cartesian argument for accepting mental-physical identity.

There are five main reductionist arguments for why we should accept mental-physical identity:[57]

(i) it is ontologically parsimonious (Feigl 1958; Smart 1959);
(ii) mental properties are functional properties which are identical to the function performing physical properties (Lewis 1966; Armstrong 1968; Kim 2005);[58]
(iii)we have grounds for accepting mental-physical identity as a Kripkean *a posteriori* identity (Block & Stalnaker 1999);
(iv)mental-physical correlations are best explained by identifying the mental and the physical (Hill 1991; McLaughlin 2001);
(v) mental causation can only be accommodated in a reductionist model.

I make no claims here about the first four of these arguments. All I am maintaining is that reductionists face the epistemic super-overdetermination problem, and will thus need an argument, such as (i) through (iv), in order to solve it.

I do make a claim about the fifth of these arguments. First, note that one of the primary motivations for being a reductionist is that it is supposed to *automatically* avoid the problem of overdetermination that non-reductionism faces *and that this is a significant strength of reductionism*. Indeed, in his assessment of contemporary reductionists' efforts to defend their view, Kim states that out of all the arguments in support of reductionism: 'The causal argument is the one [that] works best, if anything does' (2005: 124). Consider:

<u>The causal argument</u>
P1: Mental causes have physical effects only if reductionism is true.
P2: Mental causes have physical effects.
C: Reductionism is true.

[57] For a useful overview of contemporary reductionists' attempts to explain why the purported mental-physical identity relation holds, see Kim (2005: 91-148).
[58] Although Lewis and Armstrong were, like the other reductionists of their era, working with a notion of 'contingent identity' that Kim does not employ.

For as long as the overdetermination problem has been in currency, the consensus belief that NRP is the only physicalist position that faces it has been the primary support for P1. But, as we have seen, that belief is false. This undermines the main argument in support of reductionism, and thus has potentially very significant consequences for the debate between reductionists and non-reductionists.

However, there is an important caveat here. The epistemic super-overdetermination problem is an epistemic problem, not a metaphysical one. NRPists can't have metaphysically brute non-physical relations, but RPists don't face *that* problem–they say that the mental is identical to the physical, and so can allow that those identity relations are metaphysically brute. It can hardly be demanded of reductionists that it be spelled out in virtue of what metaphysical facts x is identical to itself. But what can be demanded is an answer to the question of why it should be believed that x = y, and it is this fact, more than any other, that generates the epistemic-overdetermination problem.

It seems to me that this is an interesting result. Not only have we seen how the overdetermination problem for NRP has been misarticulated, but we have seen that the basic problem has been so poorly understood that no-one has realised that a version of it is also a problem for reductive physicalism too–in direct contradiction of conventional philosophical wisdom. This state of affairs cries out for explanation. I will now offer one.

2.8 Why the overdetermination problem wasn't articulated properly

I propose four key reasons for the failure to articulate Malcolm's intuition properly. There are doubtless more, but these seemed particularly noteworthy. They are:

(i) the intuition that there just can't be pervasive overdetermination in the world;

(ii) the belief that if a purported cause is a genuine cause then no other cause can compete with it;

(iii)the failure to distinguish carefully enough between flukey, incoherent, incomplete and complete overdetermination when trying to frame the *problem*;

(iv)the inability to recognise the problem as fundamentally a problem of explaining the mental-physical necessitation relation.

I take these to be four key factors that affected the *dialectic as a whole*, rather than the work of individual contributors, as such; although, of course, examples of these factors' influence can be found in individual contributions. Let's consider each factor in turn.

I use 'pervasive overdetermination' as a term neutral between: incoherent, incomplete and complete overdetermination. This makes the meaning of the term somewhat confused, I admit, but this is simply to enable a concise way of referring to usage in the literature, which, as we have seen, is confused. Pervasive overdetermination has a bad reputation. But when the NRP-overdetermination debate began, some even questioned whether *any* unmediated overdetermination was conceptually possible (see Unger 1977; Bunzl 1979). And although such arguments did not gain much traction, a deep suspicion of overdetermination is felt by many, and the more overdetermination that's posited, the more that suspicion turns into outright hostility. More recently, the possibility that there might be pervasive overdetermination has been described as 'ugly' (Merrick 2001: 67), 'absurd' (Kim 1993: 183), 'suspiciously *ad hoc*' (Lowe 2000: 572n5), or as requiring the implausible belief that God is a 'bad engineer' (Schiffer 1987: 148).[59]

Call this general intuitive aversion to pervasive overdetermination the 'NO-OD intuition'. Some of what drove the acceptance of the NO-OD intuition has been touched on above. And if we are to explain why the NRP-overdetermination debate has been labouring with a misarticulated problem for so long, then the fact that so many in the debate shared the NO-OD intuition seems to be an important part of the puzzle. Arguments in support of the NO-OD intuition are thin on the ground, and most discussions of the problem of overdetermination supposedly faced by NRP simply run with the intuition. Pro-NRPists typically exploit the fact that it's a largely unsupported intuition and try to provide arguments as to why the intuition can be ignored, or defused; and anti-NRPists typically use the intuition as a non-negotiable premise of their arguments. Hence, for example, the debate we have seen over Kim's exclusion principle–a debate which closely mirrors a key aspect of the original exchange between Malcolm and Goldman, in which, as we have seen, Goldman questions why pervasive overdetermination should be ruled out without argument.

[59] Although using the same term, Merricks, Kim, Lowe and Schiffer were not all considering exactly the same kind of overdetermination in identical metaphysical frameworks.

I think this intuitional impasse helped to obscure the correct articulation of the overdetermination problem. Because, if the standard anti-NRP view has it that the *problem* for NRP is simply that it entails pervasive overdetermination, and that's intuitively suspect in a way not requiring much (if any) justification, then this will constrain the epistemic perspective of those anti-NRPists. Moreover, pro-NRPists will tend to meet any challenge as it is given, and so their articulations of the problem will be similarly constrained.

A further compounding factor, it seems to me, is that closely related to the NO-OD intuition is (ii): the belief that if a cause is a genuine cause then no other cause can compete with it. It is interesting to note that this belief is also the thought that motivated some of those who wished to rule out *all* unmediated overdetermination as conceptually impossible. Peter Unger, for example, in a paper *not* directed at the NRP-overdetermination dialectic, defended what he called the *principle of the uniqueness of that which causes*; as he states:

> Expressed in terms of the verb 'cause', and a notion of a *particular thing*, which delimits the relevant objects of the verb, the principle reads like this: *if some entity causes, or some entities cause, a particular thing, then nothing else causes that latter thing and no other entities cause it* (1977: 177; emphasis added).

There is a striking similarity between Unger's principle of uniqueness and the no overdetermination principle assumed by Malcolm, and explicitly employed by Kim and others. One important difference is that Kim's exclusion principle allows for flukey-overdetermination–although in a rather arbitrary way. If causes really are *unique* in something like Unger's sense, and much of what Kim says suggests similar thinking, then flukey-overdetermination is a problem case *par excellence*. And we also find something similar to Unger's thought in the recent turn in Kim's work towards productionism, which could be interpreted as an attempt to justify the belief that there's something about being a *cause* that should make us reject pervasive overdetermination.

This, I think, supports the contention that this view of *causes as unique* is another factor, closely related to the NO-OD intuition, which led to the misarticulating of Malcolm's intuition.

Then there is (iii): the failure to distinguish carefully enough between flukey, incoherent, incomplete and complete overdetermination when trying to frame the *problem*. When people think of overdetermination in relation to NRP, the starting point is invariably flukey-overdetermination: firing squads, assassins and the like. This is in many ways an intuitively appealing point of departure, but it's also a potentially problematic place to begin one's investigation from. Because of course with flukey-overdetermination there is no significant relation between the causes, so if one's thinking about the nature of overdetermination is framed too much in terms of that model, then one might miss that the unexplained necessitation relation between the two causes is what actually creates the problem for NRP. Indeed, if one starts with the idea that flukey cases are what we want to avoid, then the standard OD-Okist style move can start to look very appealing.

But what complicates matters, and makes this error difficult to spot, is that when trying to set up the overdetermination problem commentators often run-together talk of flukey-overdetermination with talk of the other kinds of overdetermination. Often, a description of what the problem of overdetermination supposedly faced by NRP is will begin with a discussion of flukey cases, but end up with a discussion of incomplete or complete overdetermination via some comments which suggest incoherent-overdetermination. This often proceeds as follows. First a paradigm case of overdetermination is considered: the flukey firing squad. Then it's suggested that, 'this would be bad if it happened in every case': i.e. there is an opaque and confusing switch to talk of incoherent-overdetermination. And, if you read everything the arguer has ever published on the subject including all their footnotes, you'll probably discover clarified somewhere that they mean either incomplete or complete overdetermination. Both sides seem more or less equally prone to this. Indeed, one might summarise the debate as follows.

If you're an *anti*-NRPist, you'll say something like, 'look at flukey-overdetermination, how mundane, but NRPists are committed to overdetermination happening in *every case*; that's a *problem*. NRPists need to tell us how they avoid that problem; if they can't then so much the worse for NRP.' For example, here's Kim:

> Genuine overdetermination is illustrated by the 'firing squad' example [in contrast, NRP] would force us to admit that every case of mental-to-physical causation is a case of causal overdetermination, one in which a physical cause, even if the mental cause had not occurred, would have brought about the physical effect. This seems like a bizarre thing to believe (2011: 216, 215).

If you're a *pro*-NRPist, then you'll say something like, 'but of course, given NRP, mental causes completely depend on physical causes which isn't at all like flukey-overdetermination that happens in every case, so problem solved.' For example, here's Bennett:

> The compatibilist could in principle accept that the effects of mental causes *are* always overdetermined, just not in a bad way–the overdetermination is perfectly acceptable, unsurprising, and unproblematic […] what the compatibilist needs to say is that the mental/physical case is importantly different from textbook examples of firing squads […] the compatibilist needs to *break the analogy* between the two types of case [and] the difference, the compatibilist will say, is that there is an important tight relation between the mental and physical that just does not hold between the two shootings (2003: 474-475).

What these examples show us is that both sides in this debate have failed to articulate the problem properly in significant part by failing to properly distinguish flukey, incoherent, incomplete and complete overdetermination *in relation to the problem they are supposed to be articulating*; and in the fray, the super-overdetermination problem has been missed.

And one of the most dramatic consequences of this miss was the embedding of the belief that reductive physicalism does not encounter any sort of overdetermination problem too. Examining this belief reveals, I think, its source: (iv) an inability to recognise the *problem* as fundamentally a problem of explaining the holding of the mental-physical necessitation relation. Once this is recognised, it quickly becomes apparent that both reductionists and non-reductionists face a problem if they are to respond to the actual challenge posed by Malcolm's original intuition: of providing a truly convincing defence of the compatibility of mental and physical causal explanation.

2.9 Conclusion

'The human body is assumed to be as complete a causal system as is a gasoline engine.' So spoke Norman Malcolm in the third sentence of the article which marks the beginning of this debate. If any one line captures Malcolm's intuition, then this is it. And we began by outlining his intuition and setting out some criteria for articulating it properly. This was to show how everyone has failed to articulate it properly, and thus missed the correct 'super-overdetermination' articulation. The two conditions laid down for a proper articulation of the intuition were: (AC1) that a genuine problem be articulated; and (AC2) that the articulation does justice to the intuition.

We then outlined what is normally taken to be the target view, NRP, and showed how the super-overdetermination articulation meets both articulation conditions. First, it's a genuine problem because NRP must be physicalist and super-overdetermination is incompatible with physicalism in virtue of the particular brute necessitation relation: dependence. Second, that brute relation also makes sense of how the two causes appear to problematically compete: in the absence of an explanation of the dependence we can't say that either cause is a winner, that both win, or it's a draw. Yet it *seems* like either one or the other *should* be the winner, or that both win, or that it's a draw. Hence the appearance of problematic competition.

The relationship between the super-overdetermination and superdupervenience problem was examined, with it being argued that the former is a special case of the latter. If so, we saw, then this has important consequences for our assessment of NRP: if NRP can't solve the former problem, then it can't solve the latter.

Next, we considered how the standard articulations of Malcolm's intuition were often expressed in a way ambiguous between four different kinds of overdetermination: flukey, incoherent, incomplete, and complete; none of which, it seemed, were *problems* for NRP. Thus, they all failed the first articulation condition, leaving the super-overdetermination problem as *the* problem of overdetermination faced by NRP.

This contention was confirmed by a detailed consideration of the best-known articulation of Malcolm's intuition: Jaegwon Kim's exclusion and supervenience arguments. Kim's exclusion argument was criticised for: (i) failing to recognise the real role dependence plays in the overdetermination problem, and instead employing the relation ineffectively; (ii) resting largely on the exclusion principle which is not 'virtually analytic' as claimed; (iii) failing to meet AC1; (iv) falling back on a productive account of causation which does not support the exclusion principle as intended; and at worst begs the question against NRP, or at best significantly reduces the force of the argument.

Penultimately, we saw an initially surprising consequence of articulating Malcom's intuition in the correct manner: that reductive physicalism faces a problem similar to that faced by NRP, namely 'epistemic-overdetermination'. This somewhat undermines one of the pillars on which contemporary reductionism stands: that NRP, and only NRP faces the task of explaining how distinct mental and physical causal explanation can be compatible. Although, as we noted, there remains an important difference between the problems faced by the two views: reductionists avoid the super-overdetermination problem because of their commitment to mental-physical identity.

Finally, we considered four factors to help explain the problems in the dialectic with properly articulating Malcolm's intuition: the NO-OD intuition; the closely related belief that causes are unique; the failure to distinguish carefully enough between flukey, incoherent, incomplete and complete overdetermination; and the inability to recognise the problem as fundamentally a problem of explaining the mental-physical necessitation relation. These four factors, it was suggested, help to make some sense of the failure to properly articulate Malcolm's intuition since its introduction into the discourse in 1968.

Chapter 3
Immanentism Assessed

Chapter Summary

In the last chapter, we examined the problem of overdetermination supposedly faced by NRP. It was argued that the traditional articulations of the problem were flawed in so far as they failed to articulate a genuine problem for NRP, and that a new articulation which does present a genuine problem for NRP was to be preferred: the super-overdetermination problem. This latter problem is a special case of the superdupervenience problem.

In this, our third and penultimate chapter, we begin by outlining a family of NRP views I call ‘immanentism’ which, although they are explicitly motivated as solutions to the traditional articulation of the overdetermination problem, *prima facie* promise a solution to the super-overdetermination problem in that they offer explanations, at least in a general sense, of the relationship between mental and physical causes. We then unpack the three explanatory criteria we first encountered in Chapter 2 (Section 2.3), against which immanentism will be assessed. This is followed by an examination of immanentism as a general approach, through the lens of the question of why Yablo and other immanentists offered epistemically distinct relations as ‘package deals’. We then consider in turn each candidate for the R required to meet the three explanatory criteria, the determinate/determinable, constituted/constituter, and set/subset relations. I argue that they all fail to meet the three criteria. This, I end the chapter by suggesting, invites a re-examination of reductionism as an alternative approach.

'Can you really be saying that mental properties stand to their physical realizations in the relation that rectangularity bears to squareness or that colors bear to their shades?' Yes. At least that is my conjecture, to be evaluated like any other by the evidence for it and by its theoretical fruitfulness.
- Stephen Yablo[60]

3.1 Immanentism: a solution in search of the super-overdetermination problem

There is a variant of NRP, best called 'immanentism', that is interesting because although immanentists articulate the overdetermination problem in the traditional way, the solution they offer to that problem looks like NRP's best bet for a solution to the super-overdetermination problem. To see why, we first need to clarify what immanentism is.

Unfortunately, there are no very detailed proposals in the literature on how to characterise immanentism. Part of the reason for this is that talk of a mental property or event being 'immanent' in a physical property or event is relatively new to the NRP-overdetermination dialectic, and there is no established usage for 'immanent' and its cognates. The term was introduced by Stephen Yablo (1992: 256-257n29; 1997: 33n22), as part of his account of what he deemed to be two immanence relations: determinate/determinable and set/subset; and he later offered a sketch of a more general account of immanence, but without using the term (2000; 2001).[61] After Yablo's first step, others discussed the same relations, often either attacking or defending Yablo's move, although apart from Haug none that I cite used the term 'immanence' (see MacDonald & MacDonald 1995; Ehring 1996; Worley 1997; McGrath 1998; Wilson 1999, 2009, 2011; Shoemaker 2001, 2007; 2013; Pereboom 2002; 2011; Bontly 2005; Funkhouser 2006; Walter 2007; Whittle 2007; Haug 2010).[62]

Yablo's usage aside, the first explicit reference to a category of immanence relations as a *type* of relation that might be appealed to in debates regarding NRP can be found in the

[60] (1992: 256).

[61] Although see his (1987), which discusses some of the general ideas he later developed further to respond to the overdetermination problem.

[62] Some might add to this list Lenny Clapp (2001) and Louise Antony (2003). However, as we have seen, they deny what I take to be the minimal interpretation of the distinctness tenet (Clapp 2001: 112-116; Antony 2003: 9), so they are not NRPists; although of course both Clapp and Antony defend the claim that the views they propose are versions of NRP.

Autumn 2008 version of David Robb and John Heil's 'Mental Causation' entry in *The Stanford Encyclopedia of Philosophy*; previous versions, starting in 2003, made no reference to immanence (see their 2013: § 'Author and Citation Information'). Robb and Heil borrowed the term from Yablo, they tell me (2013: personal communication); although they say little about what they think 'immanence' means, other than that it is roughly synonymous with 'nothing over and above', and that it is up to individual immanentists to spell out what the metaphysical details are (2013: § 6.4). This sheds little light on what the general category of immanentism is.

However, despite the somewhat nascent nature of the category, and the lack of firmly established terminology to discuss it, I agree that Yablo and those who came after him do constitute an interesting group of NRPists who are worth discussing as a distinct group, and who do share something that, to my eye, is captured well-enough with the phrase 'immanentism'. So why are they worth grouping together? It seems to me that Yablo, Wilson, Shoemaker, Pereboom and Paul all tried to respond to the overdetermination problem by offering an explanation of why dependence holds between the mental and the physical (or relevant mental and physical entities specifically). This, as we have seen, is the right strategy, and so these views are of particular interest here, and invite an assessment of whether or not the proffered explanations meet the three explanatory criteria (which I unpack in the next section):

(i) R entails dependence;
(ii) R is internal;
(iii) R holds between mental and physical events.

It should be noted that my characterisation of immanentism is a largely dialectical criterion: the view is defined by what the proponents were responding to, and how they responded. But this is not exactly a rarity in philosophical categorisation. And I do not intend to rule out that there might be interesting metaphysical similarities between all the proposals of the various immanentists, and it is these similarities which justify the label 'immanentism'. Yablo employs the determinate/determinable relation between physical and mental causes, and tries to explain the holding of it in virtue of a proper subset/set relation between the essences of mental and physical causes. Wilson and Shoemaker argue in similar fashion, except they take the proper subset/set relation to hold between causal powers 'associated with' mental and physical properties. Pereboom develops an independent account of the relation of *constitution*, but also holds that mental events are the determinables of physical events in Yablo's sense. Paul offers another constitution view, although it differs from Pereboom's in important respects and makes no appeal to the determinate/determinable relation. Thus, there are significant similarities between the specific proposals of each immanentist.[63]

However, beyond offering this observation regarding the prima facie similarities of the views, I do not wish to make any claims here about whether these metaphysical resemblances alone are sufficient to justify the grouping. The claim I do wish to make here is that, for my purposes, there is an interesting group of NRPists who share something important: they all tried to respond to the overdetermination problem by offering an explanation of dependence. When I talk of 'immanentism' I do so simply to pick out that group.

So if that's immanentism, why is it interesting? Because: although it is explicitly motivated as a response to the classic articulation of Malcolm's intuition, a person looking at the immanentist position in isolation could mistake it as a response to the super-overdetermination problem instead. This is a curious state of affairs worth examining.

Like everyone else, immanentists think the best way to frame Malcolm's intuition is in the classic manner, in particular by setting up the purported problem with a 'no overdetermination' principle of a sort similar to Kim's (see Yablo 1992: 246-248; Shoemaker 2007: 4, 13; Wilson 2011: 124-125; Pereboom 2002: 501-502; Paul 2007: 285). And, they aim to solve this purported problem by providing an avowedly non-reductionist metaphysic which they believe enables NRP to avoid the sort of bad overdetermination the 'no overdetermination' principle is supposed to rule out. If, they suggest, the mental is immanent in the physical, then this explains the dependence and

[63] Antony (1991), Marras (2003) and Bennett (2003; 2008) are also classed as immanentists by Robb and Heil (2013). Also, one might think that the MacDonalds (1986) were the first immanentists because they appealed to the determinate/determinable relation in response to concerns about the causal relevance of the mental. I don't include Bennett because she makes no attempt to explain dependence. Antony and Marras might seem to meet my definition because they employ relations of *physical explicability*, and *physical implementation*, respectively, to try to respond to concerns regarding the causal relevance of mental properties. However, I don't include them because, (i) they both focus their discussions on the problem of mental causation as faced by Davidson and so much of what they say struggles to have direct relevance for the NRP-overdetermination dialectic as understood herein; and (ii) they do not offer a very clear account of the relations they employ to try and resolve the problem. I don't include the MacDonalds because they were also focussing on Davidson and so were addressing a different problem.

helps vitiate the intuitive appearance of causal competition.[64] As we have seen, of course, the traditional articulation fails to deliver a genuine problem. Nevertheless, the immanentist solution is of great interest because of its *prima facie* relevance to what *is* a genuine problem: super-overdetermination.

Despite his failure to articulate Malcolm's intuition correctly, Yablo deserves the credit for being the first to grasp what was really at issue in the NRP-overdetermination debate. But this is revealed, ironically, not in how Yablo articulates the problem, but in how he tries to articulate the solution (this holds true for all immanentists). The best example of this can be found in what is arguably the most significant single paragraph contribution to the debate since Malcolm's original paper. Yablo suggests, in a footnote, that there is:

> a subtle interpretive question about supervenience. On the emergence interpretation, a thing's physical properties are metaphysically prior to its mental properties and bring them into being. To caricature emergentism just slightly, supervenience is a kind of 'supercausation' which improves on the original in that supercauses act immediately and metaphysically guarantee their supereffects [...] Another view is that the supervening mental properties are immanent in their physical bases; rather than giving rise to thought by some obscure metaphysical motion, certain material conditions are inherently conditions of thinking (1992: 256-257n29).

There is something noteworthy, it seems to me, in the fact that perhaps one of the most important insights into how to frame one of the most central problems in philosophy can be found in a little-remarked-upon footnote.[65] It is also noteworthy, it seems to me, that although this solution was offered to solve an incorrectly articulated problem, it is a solution which is at least of the correct type for solving the correct articulation of that problem. The reason for that is revealed in the above quote: when proposing the solution Yablo correctly identifies that what is at issue is explaining how the mental and the physical are related. And it is the placing of serious attempts to provide that explanation at the heart of immanentism that makes the view the most worthwhile version of NRP to

[64] There is much more to immanentism than this, of course. In particular, many immanentists argue that, given their metaphysic and a 'proportionality principle' that they contend holds true of causation, in many cases of supposed causal competition it is the mental cause that should be considered the *genuine* cause, not the physical. For our purposes, however, this argument can be put to one side.

[65] Jessica Wilson quotes part of the footnote (1999: 40), although primarily in relation to a critique of Terry Horgan's articulation of the superdupervenience problem.

consider when assessing whether or not contemporary NRPists can provide a solution to the super-overdetermination problem.

In addition, we can now see how immanentism promises to collapse the superdupervenience and super-overdetermination problems in the manner we discussed in chapter 2 (Section 2.4). If physical properties and events are the determinates of mental properties and events, for example, then this would solve both problems with one relation and demonstrate that the super-overdetermination problem was a special case of the superdupervenience problem.

The rest of the chapter in outline:

In Section 3.2, we detail the three explanatory criteria that NRPists must meet if they are to solve the super-overdetermination problem.

In Section 3.3, we examine the question of why Yablo and co. offered package deals in order to shed further light on the general nature of Yablo and co.'s immanentism and the differences between the current dialectic and the dialectic Yablo and co. were originally engaged with.

In Section 3.4, we consider d/d, and note some problems faced by Yablo's particular employment of the relation: that parts of his view are circular, and that to escape the circle he needs to fall back on his appeal to s/s. We then examine an argument against positing that d/d holds between the mental and the physical which rest on the possibility of fully specifying a mental cause with reference only to its mental 'dimensions' such that it is impossible for any physical cause upon which the mental cause may depend to specify the mental cause any further. I contend that this argument is sufficiently forceful for us to conclude that d/d does not hold between the mental and the physical, and so d/d fails the third explanatory criterion.

In Section 3.5, we examine a metaphysic offered by Jessica Wilson in an attempt to defend a d/d based immanentism against the objection raised in Section 3.4. I conclude that the view undermines itself, and faces a pressing 'explosion problem'. I close the section with some reflections on potential equivocations in the language used by Wilson and Yablo to claim that the mental might stand in d/d to the physical, and conclude that d/d fails the second explanatory criterion.

In Section 3.6, we begin with Pereboom's account of c/c, I object that Pereboom's view is obscurantist and possibility circular, and fails the second and third criterion. I object to Paul's view by arguing that she relies on making a brute identity claim about the relationship between mental and physical instances, and so fails the second and third criterion.

In Section 3.7, we consider s/s. I argue that Yablo, Wilson and Shoemaker's views all rely, in one way or another, on a brute identity claim between mental and physical entities. I also show how Yablo's view is fully circular, completing the criticism that I began in Section 3.4.

In concluding the chapter in Section 3.8, I suggest that these problems with immanentism invite a reconsideration of the merits of reductionism.

3.2 The three explanatory criteria

As we saw in Chapter 2 (Section 2.2 and 2.3), the super-overdetermination problem can be solved by NRP only if it can be shown that there is a relation, or set of relations, R, such that:

(i) R entails dependence;
(ii) R is internal;
(iii) R holds between mental and physical events.

Call these the 'three criteria'. The first explanatory criterion is satisfied iff, simply by stating what R's properties are, we can draw a valid inference from 'A stands in R to B' to 'A depends on B'.

We want the inference to be valid because we're after a sufficient explanation of why dependence holds; if R doesn't entail dependence in this way then it can't be a sufficient explanation of why dependence holds. So, if R must entail dependence in this way, then R must be an irreflexive, asymmetric, synchronous, non-causal, modally forceful relation, of which the following is true (where ‘A’ and ‘B’ stand for the domain of mental and physical properties, respectively):

A stands in R to B only if: necessarily, if anything x has some property F in A, then there is at least one property G in B such that x has G, and, necessarily, everything that has G has F.[66]

I leave open the question of whether or not R could, in some sense, just *be* dependence. One could try to rule this out by stipulating that R cannot be identical to dependence. Perhaps, one might think, this is required to avoid vicious circularity–we can’t explain the holding of dependence by appeal to the holding of dependence, surely? This is a plausible thought, but it misses two important points. First, it seems possible that analysing relations can avoid vicious circularity and be informative, *pace* the paradox of analysis. Examples of attempts to do this abound in philosophy: ‘causation is counterfactual dependence’, ‘explanation is law based deduction’, or, closer to home, ‘realisation is determination’. But if we are to at least allow for the possibility of an analysis like this, as it seems wise to do, then we can’t rule out that a statement of the form ‘dependence is R’ might be true.

Second, we could also explain the holding of dependence by offering an analysis of its relata which shows that it is an internal relation, and in that case we would not need a distinct R to explain the holding of dependence. Thus, building into the three criteria the requirement that dependence not be identical to R would be the wrong move. If R is internal–then no further relation needs to be appealed to in order to explain its holding, instead R’s holding can be explained by appeal to the nature of the relata.

Recall (from Section 2.3) that the sense of ‘internal’ employed here is as follows. Minimally, a relation, R, is internal to x and y iff, necessarily if x and y exist then xRy. The most obvious way to show that R is internal is to show that the natures of x and y entail that they stand in R. For example, it could be the case that the existence of Jack (x) and the

[66] Three points to note here. (i) For our purposes, we can read ‘x’ as standing for either objects or events. (ii) One might wonder why, given that dependence is transitive, transitivity is not listed among the properties that R must have. The reason is that an intransitive relation can entail a transitive one. For example, the *is the successor of* relation entails the *is greater than* relation, but the *is the successor of* relation is intransitive, the *is greater than* relation transitive. (iii) We have stipulated that R must hold between ranges of properties, but one might be concerned that not all of the candidate relations we consider stand between ranges of properties – d/d for instance. For such relations, though, treat the holding of R between the two ranges of properties as reducing to the holding of R between the relevant properties in the two ranges.

existence of Jill (y) necessitates that Jill stands to Jack in the *taller than* relation (R), and we could show this by appeal to Jill's being 6 feet tall, and Jack's being five feet 8 inches tall.[67] This brings us to the second criterion.

The second criterion is satisfied iff R is internal. This is the only way to solve the super-overdetermination problem, which is a problem of bruteness: NRPists must avoid positing either an epistemically or metaphysically brute relation between mental and physical events. Recall that a relation is epistemically brute if there is no true, informative explanation for why we should believe that the relation holds, and metaphysically brute if its holding is not necessitated by anything else. My use of 'metaphysically brute' is synonymous with what is often meant by 'fundamental'. If a relation is metaphysically brute, then it's epistemically brute; and if it's epistemically brute then we can't rule out that it's metaphysically brute. NRPists must rule out that R is metaphysically brute, because if it were then R would be fundamental–it would not be necessitated by anything else–and physicalists can only countenance physical fundamentals. As the physicalist mantra goes 'everything is physical or depends on the physical'. But R can't be physical if it's brutely trans-ordinal–i.e. if it stands between the mental and the physical without being necessitated by anything else at all. This is the superdupervenience problem we first encountered in Chapter 2 (Section 2.2): no physicalist can countenance a trans-ordinal relation which has mental entities as relata and isn't necessitated by anything else at all–let alone anything physical. Thus, R can't be brute.

So how do NRPists avoid R being brute? Three options are immediate non-starters.

First, NRPists can't simply appeal to the fact that a further trans-ordinal relation holds, R*, which entails the fact that R holds, because the problem just transfers to R*, and we have embarked on a vicious regress.

Second, NRPists can't just appeal to the fact that R depends on the physical, because this is just a more complicated version of the first option. Consider: if the fact that R holds, P, is supposed to be a physical fact because it depends on physical fact S, then we face a question: why does P depend on S? Call the fact that the dependence relation holds

[67] This conception of an internal relation is a descendent of the concept of an internal relation that goes back to the beginning of analytic philosophy and the work of Bradley, Moore and others (see Bradley 1893; Moore 1919). David Lewis also talks of 'internal relations', but in a distinct sense from that employed here (see Lewis 1983: 356n16).

between P and S ‘Q’. Note that Q is not a physical fact, it’s a fact about a relation that holds between the physical, S, and the non-physical P. So we now face a question about why Q depends on S, and we have embarked on another regress. Thus, appealing to the fact that R depends on the physical is a more complicated version of the first option.

Third, NRPists can’t claim that the fact that R holds is a physical fact, because it’s a trans-ordinal fact: it concerns what stands between the mental and the physical.

The prospects for NRP might seem pretty bleak at this point; and indeed some, such as Michael Lynch and Joshua Glasgow (2003), think that the superdupervenience problem is impossible for NRPists to solve (from which it would follow that the super-overdetermination problem is too). Ultimately, I agree with this conclusion, but before we get to that *denouement*, we need to consider a possible solution that Lynch and Glasgow don’t, which gives us our second criterion: that R is an internal relation. Indeed, nearly all of the discussion in the literature does not explicitly address the issue of whether the relation that stands between the mental and the physical is an internal relation or not. But some do take a position on the issue; for example, John Howell argues, implicitly, that mental-physical dependence is an internal relation (2009), and Andrew Melnyk explicitly defends a similar view – employing a sense of ‘internal’ close to the sense employed herein (2013: 6).

In order to further clarify this issue, let’s consider two illustrative examples of an internal mental-physical relation–one reductionist, one non-reductionist.

Reductionism provides the simplest case, so we shall begin there. Our starting question: how do we explain the holding of the relation between mental and physical events? Reductionists answer: mental events just are physical events (and similarly for properties). But here arises the worry we identified in Chapter 2: reductionists just can’t help themselves to an epistemically brute identity claim. Recall that a brute identity claim, in epistemic terms, is a claim of the form ‘A=B’ where it does not appear that A=B and no true, informative explanation can be given for why we should believe that A=B.

It is arguable that some reductionists have made something approaching an epistemically brute identity claim. As Christopher Hill has suggested:

> When J. J. C. Smart and other [reductive] materialists of the 1950s and early 1960s set out to defend their favourite doctrines, they were typically more concerned to answer objections than to construct positive supporting arguments […] Thus, instead of giving carefully formulated positive arguments, the materialists of Smart's era relied mainly on sketchy appeals to simplicity and terse complaints about the obscurity and messiness of competing views (1991: 123).

However, other reductionists have gone further in trying to avoid brute identity claims (for an overview of such efforts, see Kim 2005: 93-148). For illustrative purposes, let's consider one way this could work. In answer to the question, 'why believe that mental events are identical to physical events?' one could give the following, 'realiser functionalist' answer. First, we functionalise the target mental event type in question–i.e. define it in terms of its causes and effects, its causal role. This definition operates like a definite description: to be that mental event type just is to be some event type playing that causal role. Second, we can investigate the world to see which physical event type actually plays that causal role in each physical system type (for ease of illustration, think of 'physical system type' as synonymous with 'species').[68] Third, we can then identify the mental event type in each physical system type with the physical event type which plays the relevant causal role in each physical system type. More formally (where '*t*ME' and '*t*PE' stand for a mental and a physical event type, respectively):

The identity argument
P1: Being *t*ME $=_{\text{def.}}$ being the event type playing causal role C in system S.
P2: *t*PE (and only *t*PE) plays causal role C in S.
C: *t*ME = *t*PE.

We need not debate the virtues or vices of this sort of view–the purpose here is simply to consider what a relevant non-brute identity claim might look like, and what that might tell us about the demands of the second criterion. The key thing to note about this explanation for why *t*ME = *t*PE is that the mental-physical identity relation is internal in our sense: given the nature of the relata, x and y, it is entailed that x = y.

[68] Kim favours a view along these lines, although he says that the relevant structure types will be more fine-grained than species and will also be individuated by time (1992c: 5).

What the identity argument does, or tries to do, is to show that the relevant relata natures entail the identity–to show that we should accept that mental events are identical to physical events once we have accepted the analysis of the relata. And importantly, with relations that are entailed by their relata like this, there can be no residual questions, or no sensible ones anyway, about why the relata stand in the relation. Moreover, this way of avoiding epistemic bruteness in the case of an identity claim can be extended to mental and physical properties more generally, and thus suggests a way for reductionists to avoid the problem of making epistemically brute claims when trying to solve the broader mind-body problem.

This suggests a way for non-reductionists to solve the superdupervenience problem for events, and thereby solve the super-overdetermination problem: show that given the nature of mental and physical events, they stand in R. Whether or not we then want to accept the R in question will no longer be bedevilled by the superdupervenience problem and the attendant fears of regress; instead, we will simply face the question of whether or not mental and physical events really have those natures. And we can see how this might work with a non-reductionist analogue of the reductionist proposal we have just considered, along broadly 'role functionalist' lines, as follows.[69]

First, we again define the target mental event type in terms of its causal role–but where this is understood as the mental event type being identical to the property of being some event or other that plays the relevant causal role. This definition operates not like a definite description, as for realiser functionalism above, but instead as the name for a higher-order relational property. Second, we investigate the world to see which physical event types actually perform the causal role of the mental event type, noting that many physical event types play that role. Third, we can then see that the mental event type depends on the physical event types, because that dependence relation is entailed by the nature of the mental and physical event types. More formally:

The dependence argument

P1: Being tME $=_{\text{def.}}$ the property of being some event or other that plays causal role C.

P2: $t\text{PE}_1$, or $t\text{PE}_2$… or $t\text{PE}_n$ play causal role C.

C: tME depends on $t\text{PE}_1$, or $t\text{PE}_2$… or $t\text{PE}_n$.

[69] Melnyk has recently proposed an alternative approach which endeavours to show that mental-physical realisation is an internal relation (2013: §3).

Similarly to what we saw above with the identity argument, what the dependence argument does, or tries to do, is to show that the relevant relata natures entail the dependence–to show that we should accept that mental events depend on physical events once we have accepted the analysis of the relata. And importantly, with relations that are entailed by their relata like this, there can be no sensible residual questions about why the relata stand in the relation. Moreover, this way of avoiding epistemic bruteness in the case of a dependence claim can be extended to mental and physical properties more generally, and thus suggests a way for NRPists to avoid the problem of making epistemically brute claims when trying to solve the broader mind-body problem.

We will consider alternatives to this way of showing, on behalf of NRP, that R is internal. But first, let us consider the third and last of the explanatory criteria.

As we saw in Chapter 2 (Section 2.2), with regard to the third criterion, even if one can posit an R which entails dependence and is internal, that alone falls far short of explaining why dependence holds. One could argue, for example, that the mental stands to the physical in the relation: *is a higher biological taxa than*. The biological taxa relation entails dependence and is internal, so meets the first two criteria, but is clearly not a good candidate for R. Thus the third criterion, which stipulates that R must actually hold between the mental and the physical.

We might then ask: how do we know when the third criterion has been satisfied? It would certainly be satisfied if a sound argument were given for the holding of R between the mental and the physical. But let us simply demand an argument for the holding of R which is able to withstand any objections we might put.

Thus, those are the criteria, let us quickly remind ourselves of the candidates. Immanentists have proposed three relations as candidates for R: the determinate/determinable, set/subset, and constituent/constituted relations (henceforth: d/d, s/s, and c/c, respectively). Typically, the d/d relation has been offered in a package with either the s/s or c/c relation. As we have noted previously, Yablo posits that mental events are the determinables of the physical events on which they depend, and then explains the holding of this by arguing that the essences of mental events are a proper subset of the essences of physical events.[70] Wilson and Shoemaker argue that physical properties are the determinates of mental properties, and analyse the d/d relation as a special case of the s/s relation, which they take to hold between the causal powers 'associated with' mental and physical properties. Pereboom makes an avowedly Yabloean appeal to the holding of the d/d relation between mental and physical properties, packaged alongside an appeal to the c/c relation, which he treats, largely, as primitive. Paul is the only immanentist not to offer a package deal, holding simply that mental and physical events are bundles of property instances, differing only with respect to non-causal properties–being otherwise identical. In this sense, according to Paul, mental events stand in c/c to physical events–the former are constituted by the latter.

For our primary purpose, we need not examine these relations in their various packages. We are chiefly concerned with whether any relation is an R which allows NRPists to solve the super-overdetermination problem. Thus we need not settle the difficult issue of whether one relation is a version of, is compatible with, or can explain or analyse the other (for some views on this, see Ehring 1996; Worley 1997; Funkhouser 2006; Walter 2007; Shoemaker 2007; Wilson 2009; Haug 2010; Pereboom 2011). To see if immanentism can solve the super-overdetermination problem on behalf of NRP, all that we need to consider is whether the d/d, s/s or c/c relation, considered alone, is a good candidate for being R. This is what we shall do here. If, as I will show, none of them are plausible candidates for R, then the various theoretical packages they come in are irrelevant.

3.3 Yablo and co.'s package deals

Even though we will consider each relation posited by immanentists individually, if we first consider the question of why Yablo and co. combine d/d with s/s or c/c at all, then it will help us accomplish two important tasks.

Recall that we have given a dialectical account of immanentism above: those views which try to respond to the purported problem of overdetermination faced by NRP by trying to explain the holding of the dependence relation. But we have not said much more than that about the motivations for, and details of the position. Given that we are assessing

[70] Yablo also maintains that d/d holds between mental and physical properties, but doesn't offer the same explanation for d/d holding as he does for events. I think this is a symptom of some problems in Yablo's view, which I discuss below.

immanentism in this chapter, it makes sense to look at it more closely; and considering the question of what motivated the offering of the package deals by most immanentists will help us do that.

Second it will allow us to compare the role cast for immanentism in the dialectic herein to the role immanentists see themselves as occupying. We cannot understand what motivated immanentism without completing this second task, and assessing the package deals requires a consideration of the dialectic that immanentists were engaged in. Moreover, if we are to charitably assess immanentism then we should compare ‘our’ dialectic with ‘theirs’.

The common thread running through these two tasks is the question of why immanentists offered the package deals. So let us begin by getting one answer out of the way by granting that it is an independently interesting metaphysical question regarding what d/d is, at bottom, and whether it can be explained or analysed in terms of the s/s relation, or combined with the c/c relation. This might seem justification enough for the package deals, but if the target problem to be solved is taken to be the threat of super-overdetermination, then offering a package rather than one relation could mean that Yablo and co. are overcommitting. If d/d is just a version of s/s, then, ultimately, it is the s/s relation that is solving the problem–in the final analysis s/s would be R. In that case, one might wonder: why bother with the d/d middleman when setting out the argument? It simply saddles one’s position with more claims than is necessary; and it is generally unwise to burden one’s theories with such hostages to dialectical fortune.[71] Instead, one could simply defend the claim that s/s is R, without having also to defend the claim that d/d is, or is a version of, s/s.

So, given that the super-overdetermination problem was not anticipated by the immanentists, it will be charitable and illuminating to consider what originally motivated the offering of the package deals. And we shall do this by looking at the source: Yablo’s original proposal. Wilson and Shoemaker follow Yablo in discussing d/d and s/s together, and appear driven by similar concerns, and part of Pereboom’s proposal is avowedly Yabloean.

[71] Perhaps, as my colleague Umut Baysan has wryly observed, this is simply an instance of explanatory overdetermination.

There seem to be three main reasons for Yablo's original package deal. The first two provide an answer to the question 'why d/d at all?' the third 'why d/d and s/s (or any other relation)?' Considering these reasons will aid us in both tasks.

The first reason: Yablo is driven by what he sees as the need to respond to the traditional articulation of the overdetermination problem, at the heart of which is the NO-OD intuition. Recall that the NO-OD intuition is the rather confused thought that there just can't be pervasive overdetermination, where 'pervasive overdetermination' could mean incoherent, incomplete, or complete overdetermination. And it is the NO-OD intuition which Yablo clearly takes aim at with his initial appeals to d/d. Consider his famous example: a pigeon, Sophie, is trained to peck at all red triangles, and she is presented with a scarlet triangle, and she pecks. With the NO-OD intuition in mind, we might ask: what caused Sophie's pecking–the triangle's being red, or scarlet? Yablo contends that it would be absurd to suggest that the triangle's being scarlet caused the pecking but its being red didn't; just as it would be absurd to conclude that any determinables are causally outcompeted by their determinates. As he states:

> At last we have hit on a relation that brain states plausibly bear to mental ones […] *and* that makes nonsense of the causal competition idea. No-one would call the triangle's *redness* irrelevant to her pecking on the grounds that the effect was already provided for by its specific *shade* of red […] Nor would anyone think that my screaming as such was irrelevant since my screaming "wake up!" was sufficient. Examples like these confirm what seems obvious anyway: determinates do not pre-empt their determinables (1997: 256).

As we have seen, of course, the 'causal competition idea' can be articulated in a manner Yablo did not anticipate: in terms of super-overdetermination. Nevertheless, it helps us illuminate immanentism to note that Yablo appeals to d/d in order to disarm the NO-OD intuition upon which the traditional articulation of Malcolm's intuition rests. All immanentists believe they need to address the NO-OD intuition, and given our rejection of the NO-OD intuition as confused and unjustified, we can see here a significant difference between the dialectic that immanentists see themselves as engaged in, and the dialectic herein. We will consider specific examples of this when we assess each immanentist proposal below.

The NO-OD intuition also plays a role in Yablo's second reason for appealing to d/d. Yablo views d/d as a specificity relation, and if it is such, then d/d sits well with the standard immanentist principle that causes must be proportional to their effects, a principle appealed to in order to find a causal 'winner' and satisfy the NO-OD intuition. To elucidate, Yablo understands the d/d relation like so: a determinate, A, stands in d/d to a determinable, B, iff being A is a more specific way of being B (1992: 252). Understanding d/d either wholly or partly in terms of specificity is quite common. For example, the notion of specificity in relation to d/d has been referred to as 'fundamental' (Prior 1949: 14), the 'first criterion' (Körner & Searle 1959: 144), and as the 'central' (Funkhouser 2006: 550), or 'core' (Wilson 2009: 152) feature. If d/d is a specificity relation, then it fits neatly with the proportionality principle, which states that C is the cause of E only if C has the right 'level of detail'. Consider Sophie the pigeon again: was it the scarletness or the redness of the triangle which caused her pecking? Supporters of the proportionality principle maintain that to answer this question we must ask: is the scarletness or the redness at the right level of detail? In the case of Sophie, even though the triangle's being scarlet was sufficient to cause the pecking, claiming the scarletness as the cause is too detailed, too specific; simple redness will do–Sophie, after all, is trained to peck at all red triangles, regardless of shade.

Thus, according to the above picture, if mental causes stand in d/d to physical causes, then physical causes are simply more specific ways of being mental causes, like scarlet is a more specific way of being red. So if c-fibre activation is simply a specific way of being in pain, then all we need do if we want to know which is *the* cause of, say, the hand on the head, is to ascertain whether the c-fibre or the pain is proportionate. In this way, the notion of specificity works two jobs for Yablo and co. First in that the appeal to d/d allows us to think of the purportedly competing causes as standing in a specificity relation; which, second, opens the way to an appeal to proportionality to find a winner of that competition: the cause with the right amount of specificity. And in this way, Yablo and co. aim to satisfy the NO-OD intuition.[72]

[72] One might think that there is a tension in Yablo's view; on the hand he wants to 'make nonsense of the causal competition idea', while on the other he picks a winner of the competition. However, Yablo accepts that both the supposedly competing mental and physical causes are 'sufficient causes' of the effect, and in that sense denies that there's causal competition, whilst also holding, perfectly consistently, that the physical cause is sufficient *and then some* – it's too specific. It's the redness of the triangle that caused the pecking, not the scarletness *per se*, even though the scarletness is sufficient to cause the pecking – scarlet is just a specific way of being red, after all. Thus, on Yablo's view, there's no causal competition in any genuinely worrying way: there's just a sense in which both causes are causes – they're both sufficient – and a sense in which the mental cause is *the* cause – it's sufficient and proportional.

But this brings us back to our original concern about the need for an appeal to a package of relations. We might ask: ‘if d/d does all that, then why did immanentists bother with s/s, or any other relation?’ Thus the third reason: Yablo was in significant part driven to appeal to s/s by explanatory worries about d/d. Or, at least, I take this to be one plausible reading of his rather complex position. We will deal with this issue in more detail in the section on d/d below, but what I will say here holds at least for d/d as Yablo takes that relation to hold between mental and physical events.

Yablo expresses his concern with explaining d/d in various places; but consider, for example, when he states that:

> [if] q’s essence is a subset of p’s essence, p is said to subsume q (p ≥ q); and p [stands in d/d to] q (p > q) when the inclusion is strict […] Explaining [d/d] by essence has three points in its favour: it fits the intuitive examples; it supports the analogy [of mental and physical properties standing in d/d to each other]; and it predicts the principle that p [stands in d/d to] q only if for p to occur is for q to occur in a certain way’ (1992: 262).

It seems to me that part of the motivation for this explanatory concern is that Yablo doesn’t want d/d to be either primitive or brute. In the sense of interest here, a relation is primitive iff its nature is unanalysable.[73] For example, supervenience is not primitive: the central idea is clear, ‘cannot be an A change without a B change’, and the relation’s logical properties make sense in light of that idea. To that extent, we know what supervenience is–it isn’t a primitive relation. However, instances of supervenience can be brute in that those instances themselves aren’t necessitated by anything, or we can’t know if they are.

Of course, for any relation, questions about its nature will bear on questions about whether and why it holds. But for now, let us make clear why it would be a bad result for Yablo if d/d was either primitive or brute. What I offer here is an explication–an analysis that makes the most sense, in my view, of why Yablo avoided primitiveness and bruteness. This explication comes in three parts.

[73] When I talk of a relation being primitive, this is shorthand for the concept of the relation being primitive. If we assume the classical theory of concepts, then the claim that a concept is primitive can be understood as the claim that we cannot informatively specify the necessary and sufficient conditions of the concept. This is a useful working model to bear in mind when considering what I have to say about primitiveness, but what I say will also hold for other theories of concepts which allow for non-classical forms of conceptual analysis.

Part one: Yablo doesn't want to treat d/d as primitive in significant part because at the time of his original proposal d/d was widely considered to be a conceptual relation: if x is scarlet, then it conceptually follows that x is red. And one of the most widely agreed upon points in the philosophy of mind is that whatever the relation is between the mental and the physical, that relation is *not* conceptual: if x has c-fibres activating it doesn't conceptually follow that x is in pain; as Yablo accepted (1992: 253).[74] This meant that Yablo needed to give an analysis of d/d that justified the claim that it was not necessarily a conceptual relation, so treating it as primitive wasn't a viable option–as this would simply replace an implausibility with an opacity. Yablo addresses the 'conceptuality of d/d' issue in more than one way, but appealing to s/s at least completes his analysis of d/d such that the non-conceptual variety of d/d Yablo intends to elucidate is made clear.

And Yablo needs to complete this analysis because the main argument that he gives for why d/d need not be understood as a conceptual relation, aside from the analysis of it in terms of s/s, relies on the contention that it can be understood largely as a kind of asymmetric necessitation. But that is insufficient to make clear what this non-conceptual variety of d/d is. As Yablo states:

> just as the discovery of *a posteriori* necessities upset the traditional presumption of a conceptual equivalence condition on property identity, it also makes trouble for the conceptual entailment condition on [d/d]. Take the property of being at temperature 95° C, and some highly specific micromechanical property *K* chosen so that necessarily whatever has *K* has the temperature property, though not conversely. Since *K*s which are warmer than 95° C cannot be ruled out on *a priori* grounds alone, traditional [d/d] fails. Yet the relevance of this to the properties' strictly *metaphysical* relations is obscure; and since it is only the metaphysics that matters to causation, we should discount the traditional doctrine's conceptual component and reconceive [d/d] in wholly metaphysical terms […] So *P* [stands in d/d to] *Q* just in case the traditional relation's first, metaphysical component is in place, where this consists primarily in the fact that *P* necessitates *Q* asymmetrically (1992: 252-253, 253n23).

[74] Although see also his comments in the following footnote where he states: 'But I take it that there may also be an explanation of how thinking arises out of neurophysiology, such that if I knew it, then I would find it *in*conceivable, and consider it impossible, that something should be P without thinking. What's more, the prospect of such an explanation makes the hypothesis of an unthinking P only dubiously conceivable *today*' (1992: 254n24).

Yablo also says that it ‘probably goes too far’ to identify [d/d] with asymmetric necessitation (253n23), which invites a question: what’s the difference between asymmetric necessitation and the variety of d/d Yablo is attempting to articulate? As we will see when we return to this issue in the next section, Yablo’s only way of answering this question is his analysis of d/d in terms of s/s. Thus, Yablo can’t treat d/d as primitive because he’s trying to articulate a novel account of the relation–one that is non-conceptual. He wants to articulate that novel account in terms of a kind of asymmetric necessitation that is non-conceptual yet also not simply *mere* asymmetric necessitation. And Yablo’s appeal to s/s is his only way of properly enunciating that articulation.

So Yablo faced significant dialectical challenges in trying to justify his appeal to d/d holding between the mental and the physical, challenges he had to meet by giving an analysis of d/d. Thus treating d/d as primitive was not a viable option.

This then invites another question: what’s the difference between an internal and a conceptual relation? I think this is a difficult question to answer, and nothing I say will hang on whether or not there is a distinction. Ultimately, it is an issue for my opponents to resolve, as they are committed to the mental and the physical standing in an internal relation, and some of them cite examples of relations which are paradigm cases of conceptual relations and distinguish the relation that is a candidate for R from such conceptual relations. Nevertheless, there does seem to be an intuitive difference, and I will suggest one way of elucidating that difference. The basic thought is that to describe a relation as ‘internal’ is to make a metaphysical claim about why it holds: because the relata have certain natures. In contrast, to describe a relation as ‘conceptual’ in the sense employed above is to make a semantic claim about why one term entails another: because of the definitions of the relevant terms; or, as one might say, because of the concepts the relevant terms express. Consider the following two examples.

First: if the mass of Earth is 5.9722 x 10^{24} kg, and the mass of Jupiter is 1.8981 x 10^{27} kg, then this entails that Jupiter has a greater mass than Earth.

To claim that the *has a greater mass than* relation is internal is to make a metaphysical claim: the relation is necessitated by the nature of the relata. We haven’t here made an explicitly semantic claim about the definition of the predicates ‘is 5.9722 x 10^{24} kg’ and ‘is 1.8981 x 10^{27} kg’, or any concept they express.

Second: if x is scarlet, then this conceptually entails that x is red.

To claim that entailment from scarlet to red is conceptual is to make a semantic claim: given the definition of ‘is scarlet’ it follows ‘is red’. We haven’t here made an explicitly metaphysical claim about scarletness and redness.

There are major background issues which loom over this explanation of the internal/conceptual distinction that I will not address here. Perhaps the most obvious is the issue of whether or not there is any distinction at all between logical or conceptual necessity and metaphysical necessity (see Levine 1993; 2001; Tye 1995; Chalmers 1996; 1999; Lowe 1998; Jackson 1998; McLaughlin 2007; Kirk 2013). It is perhaps worth noting that Yablo, Wilson and Shoemaker all assume the conceptual/metaphysical distinction by endeavouring to distinguish a metaphysical version of d/d, so to that extent the above explanation of the internal/conceptual distinction sits well with regards to their views– which is a good thing because the internal/conceptual distinction is drawn largely to help make sense of them.

Part two: if the type of non-causal ontological relation that we’re concerned with here is primitive, then we have good grounds for believing that it holds brutely. Consider the explanation: *R_1 is primitive*, *and R_1 holds because of S*–where ‘S’ stands for the nature of the relata that R_1 might be entailed by. This invites a question: how do we know that R_1 holds because of S? The answer we’re working with here is that S facts entail R_1. But if R_1 is primitive how can we know if this is the case? R_1 is primitive so we have no way of knowing if the entailment goes through.

One might be tempted to respond here with what I call the ‘Goldilocks objection’. Perhaps R_1 is *partly* primitive–i.e. not so primitive that we don’t know *any* of its properties, perhaps we could know just enough of its properties to know that it is entailed by S, yet it’s still primitive. One way to respond to this would be to restrict the ‘if primitive then brute’ claim to completely primitive relations only. But we will consider a relation which is supposed to be partly primitive when we discuss Pereboom’s analysis of c/c in Section 3.5, so it will be worthwhile to consider the Goldilocks objection in more detail.

I concede that it is difficult to conclusively rule out the possibility of partly primitive relations which are not brute, but the possibility looks remote enough to leave largely unthreatened the claim that if a relation is primitive then we have good grounds for holding that it's brute. The main problem with the Goldilocks objection is that once you start positing properties of R_1 to justify its entailment by S then you are left facing difficult questions about the extent to which the relation is still primitive, and the only way to answer such questions is to make what look likely to be mystery mongering appeals to intuition. Let's unpack this.

According to the proposal, R_1 is a primitive relation yet we know just enough about it to know that it is entailed by S. But on what rests the claim that R_1 is *primitive*? We know enough about R_1 to know it is entailed by S, *ex hypothesi*, so what is it that we don't know, exactly? The only way to answer this question is to appeal to intuitions about the relation, but such appeals will have a force proportional to the extent to which the intuitions are deeply held and universally shared. And think about the cases in metaphysics where we have interesting examples of intuitions which are deeply held, and universally, or near universally shared. For example, consider causal overdetermination cases in which nearly everyone agrees that, intuitively, both overdeterminers are *causes*, in a way famously problematic for the counterfactual theory of causation. What makes such intuitions so powerful is what might be called their 'Moorean force'–their rootedness in our commonsense understanding of the world. Thus, any intuition that R_1, despite all we know about it, still contains further mysteries sufficient to deem the relation primitive, will require a high degree of Moorean force if it is to carry any weight. Given how few intuitions concerning the metaphysical nature of entities actually have such force, this seems like a remote possibility.

Thus, the Goldilocks objection is in one sense well put: for all we know perhaps there will be a relation that's just right: primitive, but not *so* primitive that bruteness follows. This is why it makes sense to claim merely that if a relation is primitive then we have good grounds for holding it to be brute. And so in another sense the Goldilocks objection is not well put: the possibility it entertains is allowed for, and should not concern us given how remote it seems.

What this shows us is that the contention we started with looks robust: if a relation of the sort we're concerned with here is primitive, then we have good grounds for holding that it’s brute. This brings us to the third part of our explication of why Yablo avoided primitivness and bruteness.

Part three: if d/d is to be R, but R is at least epistemically brute, then R cannot be shown to be an internal relation, and thus R could not solve either the super-overdetermination, or superdupervenience problem. Thus, d/d can’t be brute.

This completes our three part explication of why Yablo avoided primitiveness and bruteness. Let us pause here to summarise this section so far, before concluding it. We began with a question: ‘why the package deals?’ Answering it broke down into answering two other questions ‘why d/d at all?’ and ‘why d/d plus another relation?’ We noted two reasons for the former. First, d/d was seen as a good way to respond to the NO-OD intuition. Second, treating the relevant physical events as standing in d/d to mental events, in the sense that the former were specific ways of being the latter, sat nicely with the principle that causes must be proportional to their effects and opened up a way for picking a winner of the apparent causal competition between mental and physical events. When explaining the appeal to d/d plus another relation, I suggested that a primary motivation was to avoid d/d being primitive and brute. I then gave a three part explication of why treating d/d as primitive and brute would be a bad idea. First part: d/d was traditionally viewed as a conceptual relation, but the claim that the mental is conceptually entailed by the physical is implausible. Consequently, this makes it difficult to combine the assertion that the mental stands in d/d to the physical with the claim that d/d isn’t the traditional conceptual relation, rather a similar but distinct relation which is primitive, as it simply replaces an implausibility with an opacity. Analysing d/d as a version of s/s avoids this pitfall. Second part: if a relation of the sort we’re concerned with here is primitive, then we have good grounds for holding that it’s brute, because its primitive nature frustrates efforts to justify the claim that we can know why it holds. Third part: if d/d is primitive and thus brute, then d/d can’t be R.

All that remains now is to consider the prospects for each immanentist candidate for R, and it is towards that assessment that we now turn.

3.4 Determinates and determinables

If we are to assess d/d as a stand-alone option for solving the super-overdetermination problem, then we need to begin by saying something more about how we must view the relation if it is to stand any chance of success. First we shall consider what we must say about d/d so that it meets the first explanatory criterion, before we examine how d/d fails to meet the second and third.

Let us begin by trying to get to grips with what d/d is; it will help us to do this if we compare it to dependence. To simplify matters, let us distil dependence down to what is arguably its most important feature: asymmetric necessitation. It is not a straightforward matter to spell out just how d/d differs from asymmetric necessitation. Yablo, indeed, skirted round this issue in his initial seminal contribution to the debate, stating in a footnote that:

> For dialectical reasons, I try to remain as neutral as I can about where [d/d] leaves off and 'mere' asymmetric necessitation begins (1992: 253n23).

Before becoming a little less reticent in a later contribution in another footnote:

> As far as I can see, all that "Y is a determinate of X" needs to mean in this paper is that Y necessitates X (not because it has a metaphysically infallible way of bringing X about but) because X is immanent in or included in Y. (1997: 275n22).

Everyone has missed the significant consequences for Yablo's view that this quote reveals. Note the circularity: Yablo begins his immanentist project by trying to explain the holding of mental-physical dependence by appeal to d/d which he advertises as a way of spelling out how the mental is 'immanent in' the physical (see 1992: 256-257n29); then later, Yablo says that d/d need only be understood as asymmetric necessitation that is 'immanent'. Thus the circularity. As we shall see, the only way out of this for Yablo is if he falls back on his analysis of d/d in terms of s/s. If so, then Yablo's view, which is famous for his employment of the d/d relation, should really be better known for his employment of s/s, without which his view would be disappointingly circular.

It might be thought that a different route out of the circle for Yablo would be to make use of his sketch of a general account of immanence in terms of an intensive part/whole relation, which he offered in his most recent thoughts on this debate (2000; 2001). However, if we examine the little Yablo says on the matter, we can see that taking this route will not escape the circle. Yablo states that the intensive part/whole relation is:

> the relation in which Socrates' drinking the hemlock, say, stands to his guzzling the hemlock; or someone's driving home on a certain occasion stands to her speeding home. A is an intensive part of B iff B is what you get when A is confined to just certain possible worlds [an alternative way of putting it is]: extensive wholes exceed their parts in size, intensive wholes exceed their parts in strength […] the principal examples we have of intensive part/whole relations [are]: the relation that individual conjuncts bear to their conjunction, and the relation that determinables bear to their determinates (2000: 37-38; 2001: 67).

This is all that Yablo says to illuminate what he means by 'intensive part/whole'. That we have so little to go on may itself be reason enough to leave this aside in any assessment of Yablo's view, but in the interests of completeness let us examine this step-by-step. First, note that the Socrates and driving examples that Yablo begins with are examples of events which stand in a conceptual version of the d/d relation to each other–a version which Yablo explicitly states that he is not positing in the mental-physical case. Second, Yablo offers the definition of 'intensive part' in modal terms. But notice that his definition is simply an asymmetric necessitation claim: A is an intensive part of B iff A necessitates B but B does not necessitate A–Socrates' guzzling the hemlock (A) necessitates his drinking the hemlock (B), but not vice versa. Third, Yablo points to the two 'principal examples' of intensive part/whole relations as conjunction and d/d, but this is somewhat confusing given that elsewhere he states that d/d and conjunction should be viewed as different in kind:

> probably it goes too far to identify [d/d] with asymmetric necessitation outright; otherwise, for example, conjunctive properties [stand in d/d to] their conjuncts (1992: 253n23).

So, in summary, we have some examples of conceptual d/d, a definition in terms of asymmetric necessitation, and a confusing similarity comparison with conjunction that stands at odds with what Yablo has said elsewhere. Thus, although Yablo may have travelled some distance, it would appear that he is still inside the circle.

So why did Yablo end up back where he started? The clear answer, to my eye, is that he was too focussed on defeating the NO-OD intuition. We have already discussed, in Section 3.3, the central role played by the NO-OD intuition in the motivation and framing of Yablo's position; but, to give another example, consider when he states that:

> Admittedly, the pain/C_i: red/scarlet analogy isn't perfect ['C_i' stands for any physical dependence base for pain]. This doesn't concern me, unless the disanalogies are such as to make pain more causally competitive with C_i than colors are with their shades (1997: 275n22).

Yablo's thinking seems to be that so long as he can get his reader to see that mental and physical events needn't be causally competitive in the traditional sense, then he's achieved his central goal. However, as we have established, that central goal was the 'solving' of a pseudo-problem. The main saving grace for Yablo was that his view did at least try to explain the holding of mental-physical dependence, and thus at least promised a solution to the genuine problem of super-overdetermination. But here, we see that a major part of that explanatory effort was circular.

This leaves only two options for Yablo. The first is to fall back on his analysis of d/d as a specificity relation. As he states:

> *P* [stands in d/d to] *Q* iff: for a thing to be *P* is for it to be *Q*, not *simpliciter*, but in a specific way (1992: 252).

But the problem with this option is that Yablo analyses the notion of specificity largely in terms of asymmetric necessitation:

> *P* [stands in d/d to] *Q* (P > Q) only if: (i) necessarily, for all *x*, if *x* has *P* then *x* has *Q*; and (ii) possibly, for some *x*, *x* has *Q* but lacks *P* (ibid).

Yablo leaves room for other conditions, but if they can't be spelled out then circularity threatens again: d/d would be defined as asymmetric necessitation plus we know not what.

Thus, even if one is prepared to accept that ‘mere’ asymmetric necessitation can be an internal relation and thus would not need its holding to be explained via an analysis of the relation, at the very least, for all his explanatory work, Yablo appears to have made little progress. The only remaining resources that Yablo has in his view that we haven’t already ruled out is his analysis of d/d in terms of s/s. This is Yablo’s second, and sole remaining option. As we will see, Yablo’s appeal to s/s ultimately provides no way out either, but that conclusion will be postponed until the section on s/s below.

But even if we grant that the appeal to s/s might be a way out of the circle for Yablo, this leads to a further problem: if Yablo were to fall back on s/s, then the fact that he only appeals to s/s to explain the holding of d/d between events and not properties, means he would be left with no explanation of d/d holding between properties. To explain: Yablo begins his seminal ‘Mental Causation’ paper by stating that there is little difference whether one frames the overdetermination problem purportedly faced by NRP in terms of competing properties or events (1992: 247). Then, for the first half of that paper he discusses d/d as it stands between properties only, analysing it as a specificity relation which asymmetrically necessitates in a way which, in Yablo’s view, could account for the holding of mental-physical dependence (1992: 250-260). In the second half, Yablo moves onto his account of the holding of d/d between events, and only there does he brings in s/s– maintaining that mental events stand in d/d to physical events in the sense that mental event essences are a proper subset of physical event essences (1992: 260-273). Importantly, Yablo does not say the same thing about properties–e.g. that mental property essences (whatever they might be), are a proper subset of physical property essences (ditto). This is why Yablo can only appeal to s/s standing between events, but not properties, in order to escape the circle. I don’t see an easy way out of this for Yablo without making significant changes to his view.

I think that these ad hominem points are interesting in themselves, but for our purposes we should pay most attention to the difficulties Yablo faced in trying to avoid simply treating d/d as asymmetric necessitation. The general lesson is that it’s just not easy to say what non-conceptual d/d is over and above the properties it shares with (metaphysical) dependence.

Thus, there are significant challenges in the way of analysing the nature of d/d; and I don’t propose to try to meet those challenges here (for an overview of a century of effort in that regard, see Sanford 2014). We're concerned with the prospects for d/d being R. I think those prospects look dim if, as we have seen, the arch-proponent of d/d as the solution to the overdetermination problem struggled to offer a particularly clear account of what d/d is, other than by awkwardly analysing d/d in terms of s/s only when d/d holds between events, and thus leaving his account of d/d as it holds between properties rather twisting in the wind. This all adds to my case, but let us be charitable. Other than the analysis in terms of s/s (which we shall assess below), the one, relatively clear thing that Yablo and others say about d/d over and above asymmetric necessitation is that it is a 'specificity relation'. So I suggest that we set the bar low in four ways.

First by treating the specificity condition as necessary for d/d: A is a determinate of B only if A is a specific way of being B.

Second, by simply granting that said condition is not just an asymmetric necessitation claim by another name.

Third, by postponing the worry (to be returned to in Chapter 4, Section 4.1), that if the mental stands in d/d to the physical, then it licences a move towards reductionism of the disjunctive physicalist sort, in the manner we discussed in Chapter 1 (Section 1.5).

Fourth, by not immediately ruling out that *is a specific way of being* (henceforth: *specifies*) can be internal and non-conceptual. Conceptual d/d is clearly internal: the fact that x is red and the fact that x is scarlet together entail that the latter specifies the former. But Yablo is explicitly trying to articulate a non-conceptual form of d/d–metaphysical d/d. Perhaps, let us grant, such a relation might be non-conceptual and internal. For example, if x is a c-fibre activation, it doesn't conceptually follow that x is a pain. Nevertheless, perhaps given the nature of the relata, *c-fibre activation* and *pain*, that the former specifies the latter is entailed. We will return to this issue below, where I will argue that there are grounds for thinking that this isn't possible because *specifies* seems to be a conceptual relation, but first let's just see if the proposal that d/d is R can fly.

As will now be shown, the proposal can't fly: d/d cannot be R. This is because physical causes are not specific ways of being mental causes, and thus d/d, understood in terms of *specifies*, fails the third explanatory criterion.

To demonstrate this, we first need to unpack the notion of specificity a little more, focussing on one event being a more specific version of another. One principle that we may take to govern the notion of specificity is:

The qua principle: if A specifies B, then A specifies B only in B-ways.[75]

To use the standard example, a specific shade event (A) specifies a colour event (B) only in respect of hue, saturation, and brilliance (B-ways). Call these B-ways 'B-dimensions'.[76] Thus: hue, saturation, and brilliance are the B-dimensions of colour events.[77] As we will see below, one might argue that there are other dimensions of colour events, but grant these for the moment. Also note the apparent conceptual relation between the definitions of 'is red' and 'is scarlet'.

Now consider the following example of a mental event which intuitively appears to have exclusively mental dimensions. To believe something, it might seem, one must only hold a certain *content* with a certain *confidence*: imagine two people who believe with absolute certainty that 2=2; in this instance it looks true to say that necessarily they have identical beliefs which cannot be further specified. Thus, if *being a content* and *being a degree of confidence* are mental properties, then it looks as if referring to mental properties can at least be sufficient for specifying belief events. And perhaps the point generalises: take any occurrence of a belief, and to fully specify it you need only refer to properties along the content and confidence dimensions. And again, note the apparent conceptual relation between the definitions of 'is content x with confidence y', and 'is a belief'.

Can we explain the appearance of being able to fully specify beliefs along content and confidence dimensions by appeal to the qua principle? There does seem to be something intuitively appealing in the claim that one can specify a belief *qua* belief only by reference to appropriate *belief-properties*. And, if we reflect on the fact that we would normally count the appropriate belief-properties of content and confidence as mental, not physical,

[75] I borrow the term 'qua principle' from Wilson (2009: 156), who offers a similar principle about d/d. W. E. Johnson makes comments suggesting he might support a principle like this about d/d (1921: 175), as does Prior (1949: 13), Körner & Searle (1959: 143), Funkhouser (2006: 550), and Walter (2007: 220).

[76] See Funkhouser's 'determination dimensions' (2006: 551).

[77] There are different ways that one might understand what a dimension is (see Funkhouser 2006: 554-557; Wilson 2009: 165-168). One way that I find useful is to think of a dimension of an event (or other particular) as a range of properties. Thus, two events can differ 'along a dimension' by each event having a different property from the range which makes up the dimension. Two colours, for example, can differ in *hue* (along the hue-dimension) by having a different hue property.

and if we grant that the case of belief generalises to all mental events, then there is apparent force in the claim that mental events can *only* be specified by reference to mental dimensions. This would also sit well with the appearance of the conceptual relation we noted above between the definitions of the content and confidence, and belief predicates.

We will consider an attempt below to deny that mental events can be specified only by reference to mental dimensions, but here I am simply pointing out that once we recognise the force of these claims:

1. If A specifies B, then A specifies B only in B-ways.
2. If scarlet specifies red, then scarlet specifies red only in red-ways.

Then, in the case of mental events, we should recognise that the intuitive, default view is that the ‘B-ways’ must be mental. Moreover, I am highlighting that, in all of the cases of plausible specification that we have considered, there appears to be a conceptual connection between, on the one hand, the definitions of the specifier predicates, and on the other, the specified’s predicate.

Thus, the intuitive, default view is that one can specify mental events if, and only if, one makes reference to their mental properties, and this seems to make sense because, when you consider plausible cases of specification, they involve a conceptual connection between the definitions of the specifier and the specified’s predicates. Such a conceptual connection would be extremely surprising if the specifier were physical and the specified mental because, of course, it seems unlikely that the mental stands in such a conceptual connection to the physical.

This should motivate us to see if we might object to the claim that the physical events which mental events depend on are specific ways of being those mental events. And we can do this without simply insisting that *specifies* is a conceptual relation (although we shall return to this thought below). Recall the qua principle and consider a case where a mental event, ME, has been fully specified along its mental dimensions, and depends on a physical event, PE. Think of the belief example above if it helps, and consider:

The exact argument[78]

P1: If ME can only be fully specified along its mental dimensions, then PE can't specify ME further.
P2: ME can only be fully specified along its mental dimensions.
C: PE can't specify ME further.

Call an event that is fully specified 'exact'. It seems relatively straightforward to fully specify any mental event along intuitively plausible mental dimensions such that it is exact. Given the exact argument, it looks like game over for d/d, understood in terms of specificity, as a candidate for R.

And it is important to note that, given the dialectic, the claim that d/d is not R can withstand a significant number of counterexamples to the exact argument. Because even if there are some mental events which can't be made exact only by reference to their mental dimensions, this would have to be true for all mental events if d/d is to be R.

But we should see if there is a quick way to respond on behalf of d/d here. One might try to deny the first premise as follows. Consider an occurrence of the certain belief that 2=2. Now consider the physical event it depends on; a question: does the physical event have mental specification dimensions? If the answer is no, it really is game over for d/d as a candidate for R, so let's grant that it does. The physical event's mental specification dimensions can either be identical to the mental event's mental specification dimensions, or not. If they're identical then the exact argument stands–the physical event wouldn't *further* specify the mental. So let's say they're not identical, how can these physical dimensions further specify a mental event which is already fully specified along its mental dimensions? We will consider an attempt to answer this question shortly, but for now just note that it is not obvious how such a question could be answered. So, the exact argument resists easy refutation.

The exact argument relies on the qua principle, which is embodied in P2. We will now examine an attempt to deny P2.

[78] The 'exact argument' is my name, although Ehring was the first to make the basic point (1996: 473). Similar thinking can be found in Funkhouser (2006: 563-564), and Walter (2007: 229-231). One important difference between Ehring and co.'s way of putting the point and my own is that the former frame the issue in terms of an essential difference between d/d and realisation, such that if A realises B, then A can't be a determinate of B. In putting forward the exact argument, I do not rely on any claim about realisation, nor any claim about dependence labelled as 'realisation'.

3.5 Different schemes, different dimensions

Jessica Wilson has recently tried to respond to the exact argument by offering a metaphysic which is intended to demonstrate how P2 might be challenged. In short, Wilson argues for a 'hybrid view' whereby mental events can be specified along both mental and physical dimensions, and what it is to 'fully specify' an event will depend on the scientific scheme in which the specification is occurring.

I contest Wilson's argument on two grounds. First, Wilson's own model assumes that mental dimensions are required in every scientific scheme for specifying mental events and physical dimensions are not, in a way which shows us that Wilson has failed to establish the required conclusion that P2 of the exact argument is false. Second, if Wilson's reasons for holding that mental events can be specified along physical dimensions are granted, then the specification dimensions of mental events explode absurdly in size. We will close this section by considering a possible defence of Wilson whereby her view is amended to hold that mental events can be specified only along physical dimensions. As we will see, the explanation for why this defence fails is instructive for why Wilson's hybrid view also fails. I offer three reasons as a diagnosis, which will also show us why d/d fails the second and third criteria. We shall also see that Wilson inherits two of Yablo's problems: focussing on the NO-OD intuition, and treating d/d as a kind of asymmetric necessitation.

Wilson frames her argument in terms of properties; in my reconstruction below I frame matters in terms of events in order to allow the dialectic to follow on more neatly from the above discussion of d/d. Nothing will hang on reframing the debate in terms of events. Wilson's argument proceeds as follows.

In order to challenge the claim that mental events are fully specified *simpliciter* by their mental dimensions, Wilson appeals to some examples which are supposed to show that even if a given event is fully specified in a given scientific scheme, that is not to say that it is fully specified *simpliciter*, since there are other ways to specify the event in other scientific schemes. To try to show that this is possible, Wilson appeals to metameric colour cases. It was asserted above that colour events have the dimensions of *hue*, *saturation* and *brilliance* (henceforth: tristimulus values). But this is not always so, claims Wilson, pointing to the phenomenon of metamerism. Colours are metameric for an observer if they are indistinguishable to that observer and yet differ in spectral composition; colours can be metameric for one observer, but not for another, depending on differences in observers and viewing conditions (Hardin 1988: 64; Maund 2012: §6.3). Wilson notes that understanding metamerism involves appeal to the physical features of the relevant events and light–in particular spectral power distributions (SPDs): the proportion of total light emitted at every visible wavelength (2009: 162). What is uncontroversial is that SPDs seem indispensable when trying to explain metamerism. For example, to simplify, it might be said that x and y match metamerically because there is a spectral power distribution variance which, in these particular observer and viewing conditions, is not detectable by the observer, but is detectable in other observer and viewing conditions.

These explanatory considerations are central to Wilson's case, but Wilson also puts forward two more general considerations to support treating SPDs as forming part of the specification dimensions of colour events:

(W1) Unifying explanatory value in philosophical theory.

(W2) The treatment of metamerism in colour science.

We will examine below the plausibility of the appeal to (W1) and (W2), but first note the two morals that Wilson claims can be drawn from treating SPDs as part of the specification dimensions of colour events. First, if colour events are mental, and if SPDs are part of the specification dimensions of colour events, then:

Metameric moral 1: mental events may have explicitly physical specification dimensions (2009: 163).[79]

Second, if in the scheme of normal-appearance colour science tristimulus values exhaust the dimensions of colour events, and if in the scheme of metameric colour science the dimensions of colour events are not exhausted by tristimulus values, then:

Metameric moral 2: different sciences may treat the same event as having different specification dimensions (ibid).[80]

[79] In Wilson's words: 'Psychological determinables may have explicitly physical determination dimensions.'

So Wilson's general contention is that the exactness of mental events need not be exactness *simpliciter*. Rather, there can be mental events fully specified along mental dimensions only in the scheme of psychological science–so long as there is a lower-level science any event can be further specified, according to the relevant scientific scheme. Hence, claims Wilson, the exact argument is unsound.[81] If Wilson is right, the claim that the mental might stand in d/d to the physical would remain standing. However, as I will now show, we should reject Wilson's position.

Let us examine Wilson's appeal to philosophical and scientific considerations in turn:

(W1): Unifying explanatory value in philosophical theory.

Wilson does not go to great lengths to explain what she means by (W1). But it seems that Wilson believes that granting colour events partly physical specification dimensions opens up the way for treating as non-exclusive the subjectivist 'inside the head' versus the objectivist 'outside the head' metaphysical options for the correct analysis of colour. At one level of grain colour events *are* inside the head–when specified along their tristimulus value dimensions–at another level of grain colours *are* outside the head–when specified along physical, SPD dimensions (2009: 161-162).

Given the little that Wilson says about (W1), it is difficult to assess. But we need not let that detain us–I'm happy to grant that, if Wilson's right, an *entente cordiale* can be achieved in the philosophy of colour. That seems true of the possibility for 'unification' in any area of enquiry. But that doesn't make any claims about unification true. We want to know what the facts are, surely, be they cordial or not. So let's move on to Wilson's second appeal:

(W2): The treatment of metamerism in colour science.

[80] In Wilson's words: 'Different sciences may treat the same determinable as having different determination dimensions.'

[81] It should be noted that Wilson was considering Ehring, Funkhouser and Walter's versions of the exact argument, which differ in significant ways (see, respectively, 1996: 472-473; 2006: 563-565; 2007: 229-31). However, Wilson's arguments, if they work at all, would work against the exact argument as I state it, hence the consideration of them in the current dialectic.

Wilson contends that scientists colloquially speak of metamerically matching colours as ‘kinds of colour’. Second, Wilson points out that in the formulation of generalisations and laws colour science is concerned with SPDs as well as tristimulus values (henceforth: SPDs are *nomically important*). Thus, maintains Wilson, these ways in which colour scientists treat metamers suggests that SPD properties could form part of the dimensions of colour events. And, thinks Wilson, the point generalises:

> there is no in-principle problem with mental [events] having explicitly physical [specification] dimensions. Moreover, this possibility is arguably actually realized, since some sciences treat mental [events] at a level of grain sensitive to explicitly physical [specification] dimensions: neuroscience, cognitive science, psychopharmacology, human and animal physiology, etc. So, for example, in psychopharmacology different forms of depression depend on whether the depression results from disorders in serotonin-based neuronal circuits or dopamine-based neuronal circuits; depending on which transmitter is involved, different drugs are likely to produce a beneficial effect (2009: 164).

This is how Wilson challenges P2 of the exact argument: mental events can have physical specification dimensions in the schemes of the relevant sciences, and because events have different specification dimensions according to different scientific schemes, events specified in scientific scheme A can always further specify events which are ‘fully specified’ in scientific scheme B, even granting that the event in question is fully specified in scientific scheme B.

I think there are two major problems with Wilson’s position. Let’s consider each in turn.

First, Wilson’s argument is undermined by her own examples, and the model that subsumes them. Consider Wilson’s examples and note that her claim that we should expand the specification dimensions of mental events to include physical specification dimensions relies on the proposed extra physical dimensions being nomically important for the mental dimensions.[82] In addition, when discussing the examples, Wilson contends that different scientific schemes may treat the ‘same’ event as having ‘different specification

[82] Putting to one side Wilson’s suggestion that colour scientists speak colloquially about metamerically matching colours as ‘kinds of colour’. Wilson offers very limited evidence of this (2009: 162n17), but even if true it does little to establish the point at issue. Scientists, like the rest of us, will often speak colloquially in ways that may, or may not be metaphysically correct; if one thinks that those ways of talking are correct, then one should provide arguments to that affect–particularly if the issue is a matter of controversy.

dimensions'–this is the second metameric moral. This suggests the following model. Consider a mental event: ME. Grant that ME has the following mental specification dimensions: A, B, C. Wilson's argument can only work if we should hold that: for any scientific scheme, S, for all x in S, x is part of the specification dimensions of ME iff x is nomically important for A, or B, or C.[83] Thus:

Level	Dimensions of ME					
Psychology	A	B	C			
Science 1	A	B	C	D	E	F
Science 2	A	B	C	G	H	I

The first thing to notice about this model is that it shows us that Wilson's own examples pre-suppose that the mental specification dimensions are required in every scientific scheme; the physical specification dimensions are not. Consider Wilson's main example: colour. She tells us that SPDs should form a part of the specification dimensions of colour events because SPDs are nomically important in colour science–demonstrated by the fact that we can explain metameric effects only if we appeal to SPDs. But what that really shows is that SPDs have the status they do in colour science *only because of their nomic importance for tristimulus values*. If SPDs were not nomically important for tristimulus values, we wouldn't be talking about them. This contrasts with the fact that tristimulus values are not considered part of the specification dimensions of colour because of their nomic importance for SPDs, or anything else. But, *according to Wilson's own model*, for any of the purported physical specification dimensions of a mental event, those physical dimensions need to be nomically important for the mental specification dimensions to even enter into consideration. Thus, on Wilson's model, the mental specification dimensions are required in every scientific scheme–colour can't be specified without them–but all of the purported physical specification dimensions are not so required–there will always be some scheme in which the purported physical specification dimensions of this or that science are not required to specify colour events.

Consequently, we might agree with Wilson that 'Colour science is not concerned only with colours as individuated by hue, saturation, and brilliance' (2009: 162); but add: 'of course, it is only in so far as other properties have nomic importance for hue, saturation, and brilliance that they are of any interest.' Wilson cannot disagree with this caveat given that

[83] Grant that A, B, and C are (trivially) nomically important for themselves.

her main example assumes it; and the point carries over to the other examples Wilson cites. For instance, her claim that ‘in psychopharmacology different forms of depression depend on whether the depression results from disorders in serotonin-based neuronal circuits or dopamine-based neuronal circuits’ (2009: 164), assumes the caveat ‘of course, it is only in so far as the other physical properties have nomic importance for depression at the mental level that they are of any interest.’

And here’s the rub: if mental dimensions are required in every scheme but physical dimensions are not, then why say that the latter *specify* mental events at all? Remember the dialectic: we began with the default view that mental events can be fully specified only by reference to their mental dimensions. Wilson agrees that this is the default view (2009: 159, 161-163), and so must agree that her position faces the burden of argument. In order to try and meet that burden, Wilson tries to point to examples where it is plausible to hold that physical dimensions can specify mental events in addition to the way in which those events are specified by reference to their mental dimensions. But upon examination of those examples, we see that Wilson’s argument rests on pointing to the asymmetric nomic importance of physical dimensions for the mental dimensions in a way that inadvertently demonstrates that mental dimensions are the only dimensions required in every scheme to specify the mental events. It follows from this that, on Wilson’s model, you can’t specify a mental event without reference to its mental properties, but you can specify a mental event without reference to its physical properties. Moreover, the asymmetric nomic importance of physical dimensions for the mental dimensions can be accepted without it following that those physical dimensions *specify* the relevant mental events in any sense sufficient to contradict P2 of the exact argument. Thus, Wilson has simply failed to establish the required conclusion.[84]

And let us not forget that the exact argument can withstand a significant number of counterexamples. Because even if there are some mental events which can’t be fully specified by reference to their mental dimensions, this would have to be true for all mental events if d/d is to be R. And we have just seen how Wilson has failed to show that this is true for *any* mental events. And not only has Wilson not proven her point, her own model provides the material for a counter-argument that Wilson’s view, if accepted, leads to absurdity. This brings us to the second major problem faced by Wilson’s position.

[84] It should be noted that Wilson is largely focussed on a different dialectic concerning whether realisation essentially differs from d/d. I make no claim here about how her arguments fare in that dialectic.

Second, if Wilson has failed to establish the conclusion that physical dimensions can *specify* mental events, then we can ask: why promote SPD properties from simply being nomically important for certain tristimulus values in a manner that colour scientists understandably want to analyse and manipulate to being considered part of the specification dimensions of colour events? Or why conclude, as Wilson puts it, that 'colours of the same hue, saturation and brilliance can differ *in respect of colour*' (2009: 163; my emphasis)? Consider a physical dimension, P, and a mental dimension, M, the latter of which is had by all mental events of type E*. The implicit inference that Wilson seems to be working with is from *P is nomically important for M* to *P is a part of E*'s specification dimensions*. But this proves too much. Return to the colour example: a very wide variety of different physical dimensions are nomically important for tristimulus values: material composition properties, SPD properties, human eye properties, visual cortex properties (see Hardin 1986: 1-52; Maund 2012: §4). And that's just some of the more obvious properties at the macro-level. What about all the micro-physical properties that are nomically important for tristimulus values? The mind begins to boggle. Thus, if we accept Wilson's implicit suggestion that we should accept all physical dimensions which are nomically important for the relevant mental dimensions of colour events, then the specification dimensions of those events explode rather absurdly.

Absurdly–really? Yes–the absurdity rests on the fact that we are supposed to be *specifying* events. But if 'A specifies B' directly tracks 'A is nomically important for B' in the way Wilson assumes, then what is left of the sense that 'A *specifies* B'? Intuitively, if A is a more specific way of being B, then A is just one way of being B, like scarlet is just one way of being red. Nomic importance just seems to be something else entirely. Now, perhaps that thought is wrong, but Wilson has not given us grounds to reject it–she has simply ignored it. And in doing so, Wilson has asked us to accept something that is intuitively absurd, that mental events can be *specified* by a massive range of physical properties so long as those properties are nomically important for the relevant mental properties of those events.

Thus, Wilson's own examples, and the general model that subsumes them, undermine her argument. According to Wilson's model, mental dimensions are always required to specify mental events, physical dimensions are not. All Wilson shows is that physical dimensions can be nomically important for mental dimensions. But it does not follow from this that physical dimensions can specify a mental event. And the inference required to move from nomic importance to inclusion in the specification dimensions is too strong; if valid then the specification dimensions of mental events explode absurdly in size in a manner which does violence to the ordinary notion of specification–and Wilson does not give us good reason to commit such violence. In this way, not only has Wilson failed to establish the required conclusion, her own model shows that the required conclusion should be rejected. This leaves the default view, that mental events, qua mental events, have only mental specification dimensions, not only unrefuted, but strengthened.

This invites a question: why did Wilson's argument fail in these ways? I offer the following three reasons as a diagnosis, which will also show us why d/d fails the second criterion in addition to, as we have already witnessed, the third.

The first reason is that Wilson's defence of d/d struggled under the strain of being a hybrid proposal–it was in trying to justify why mental events should be specified along physical as well as mental dimensions that Wilson came unstuck. So why didn't Wilson just go all in and defend a 'pure view'–that mental events can be specified along only physical dimensions? The problem with this option is that it would entail that all events are *purely* physical events–on the assumption that if an event can be specified along only physical dimensions, then it is a purely physical event. This would significantly undermine the claim that there are mental events. For sure, there might be events which had mental properties not required to specify them, and in that sense might be considered mental, but it would be only in a rather attenuated sense, far short of the requirements of NRP. And seeing this can help explain why Wilson's proposal encountered the difficulties we detailed. If mental events have specification dimensions at all, then they *must* have mental specification dimensions, otherwise they aren't mental in any sense robust enough for NRP. In this way, mental dimensions come 'for free' if you're signed up to NRP: you don't need to justify their inclusion. Physical specification dimensions aren't like that–if they come at all, it's at a cost: you need to justify their inclusion. We can see that cost most clearly when considering the pure view, which no NRPist can afford to adopt. Wilson's hybrid proposal mitigates the first major problem with the pure view–that mental events are reduced to, or eliminated in favour of physical events–by not dispensing with mental dimensions. But Wilson fails to deal with the second major problem of the pure view: that physical specification dimensions need to be justified. And, it was in failing to provide a plausible justification, indeed, providing grounds for a plausible counter-justification, that Wilson's defence of d/d as a candidate for R fell short.

The second reason is one we have seen before, and will see again: the NO-OD intuition. The main reason Wilson offered a d/d specification-based metaphysic was in an effort to try and defeat the NO-OD intuition. And Wilson's effort falls into something of a general pattern. In trying to defeat the NO-OD intuition, immanentists have proposed metaphysics which allow them to say 'here is an explanation of how you can have "two causes" that do not "compete" in the problematic NO-OD intuition sense, thus NRP doesn't entail problematic overdetermination.' Consider an example each from Yablo and Wilson:

> At last we have hit on a relation that brain states plausibly bear to mental ones […] *and* that *makes nonsense of the causal competition idea*. (Yabo 1997: 256; emphasis added).

> determinables and determinates aren't causal competitors. Intuitively, *Increased specificity* characterizes a relation between properties that is sufficiently intimate to prevent overdetermination of the firing squad variety, in which case causal exclusion of either property is unmotivated (Wilson 2009: 153).

These quotes are examples of the extent to which the metaphysics offered by immanentists were motivated by the perceived need to deal with the NO-OD intuition. However, as we have seen, understanding the overdetermination problem in the traditional, NO-OD intuition sense was a misunderstanding. And once the dialectic is correctly reframed in terms of super-overdetermination, and the metaphysics are assessed solely as plausible candidates for R, then the cracks really start to show. And, as we have just seen, the metaphysic Wilson offers is no exception to this.

The third reason is that, apart from the analysis of d/d in terms of s/s which we will assess below, Wilson follows Yablo in often seeming to understand specification largely in terms of asymmetric necessitation, even if that understanding is not always made explicit (for example, see 2009: 152-153, and *passim*). And if one proceeds with such an understanding, then the hypothesis that physical events specify mental events seems far more plausible–about as plausible, indeed, as the hypothesis that physical events asymmetrically necessitate mental events. Moreover, if by 'specifies' all that is really meant is ‘asymmetrically necessitates’, then the exact argument loses its purchase.[85] Thus, if Wilson conceives of specification largely in terms of necessitation, then this would explain both why she was attracted to the view she defends, and why, by my lights, her account fails to do justice to the notion of specification–which seems to be a much meatier notion than, as Yablo put it: ‘“mere” asymmetric necessitation’ (1992: 253n23).

How much meatier? This is a difficult question to answer quickly, but the default view is surely that *specifies* is a conceptual relation. Returning to the thought we introduced above, if you consider the prima facie plausible examples of *x specifies y* that we have looked at, they were all examples of a conceptual relation. We then looked at attempts to understand the *specifies* relation in a non-conceptual way and saw how those attempts failed, even in the absence of any positive attempt to defend the claim that *specifies* is a conceptual relation. Thus, we have the outline of a possible explanation for the failure of Yablo and Wilson’s appeal to d/d understood in terms of specification. If for x to specify y requires that x and y stand in a conceptual relation, and if the mental doesn’t stand in a conceptual relation to the physical, then the failure of Yablo and Wilson’s views was to be expected.

Penultimately, I wish to make a related point about the potential pitfalls of ambiguous language. Another way in which Wilson follows Yablo is in often referring to d/d by the term ‘determination’. In other dialectics, saying that A determines B is often treated as the equivalent of saying that A necessitates B. Note, though, that there is more prima facie plausibility in the claim ‘physical events *determine* mental events’ than the claim ‘physical events *specify* mental events’. And in many of Yablo and Wilson’s passages where ‘determines’ is used with the intention of referring to d/d, it is easy to read ‘determines’ in the ‘necessitates’ rather the ‘specifies’ sense. Thus, one might explain the failure of Wilson and Yablo’s arguments as in part stemming from a problem of equivocation. Consider: Yablo and Wilson begin by considering ‘classic’ conceptual d/d–red/scarlet, drinking/guzzling, and so on. But Yablo and Wilson claim that they have identified a version of d/d that is non-conceptual–although they often continue to use the conceptual examples of d/d throughout their arguments. Apart from the appeal to s/s, when pressed on what that non-conceptual variety of d/d is, Yablo and Wilson offer an intuitive notion of specification that is either not analysed, or understood as asymmetric necessitation plus we know not what. All the while, these claims are dressed (or perhaps camouflaged) in the

[85] What also then loses its purchase, though, is the force of the original appeal to d/d which was, in Yablo’s words, supposed to make ‘nonsense of the causal competition idea’ (1997: 256). Asymmetric necessitation does not quite have the same decisive intuitive effect.

language of ‘determination’. So, on the one hand, Yablo and Wilson want to make ‘nonsense of the causal competition idea’ (Yablo 1997: 256), by appeal to the absurdity of claiming that there can be causal competition between red and scarlet, and on the other hand want to make sense of claiming that the mental stands in a ‘metaphysical’ version of d/d they call ‘determination’. But note the tension. First, ‘A necessitates B’ does not as obviously make nonsense of A-B (NO-OD type) causal competition as ‘A specifies B’ does. Second, ‘physical events are a specific way of being mental events’ seems far less plausible than ‘physical events necessitate mental events’. And it seems to me that this tension is often disguised by the use of ‘determination’ in a sense which is not always clear, and sometimes seems to shift, as the dialectic requires, between ‘specification’ and ‘necessitation’. This is matched by a tendency to use phrases like ‘is a determinate of’ in manner which seems difficult to distinguish from ‘asymmetrically necessitates’.

Examples of this can be found throughout Yablo and Wilson’s contributions to this debate, but I think the following passages are as good examples as any. First, note how Yablo easily switches between talk of d/d and talk of necessitation:

> Just as pain is not identical to any of the brain states C_i that necessitate it, red is not identical to any of the more precise shades R_i (scarlet, crimson, etc.) that necessitate *it*. Yet there is no question of redness *depending* on scarlet, for to be scarlet is *already* to be red. Scarlet is, as we say, a *way* of being red, or, in an older terminology, a *determinate* of redness. Why shouldn’t the C_is likewise be determinates of pain? [Yablo then states in the attached footnote:] As far as I can see, all that “*Y* is a determinate of *X*” needs to mean in this paper is that *Y* necessitates *X* (not because it has a metaphysically infallible way of bringing *X* about but) because *X* is immanent in or included in *Y* (1997: 256, 256n22).

Second, note Wilson’s use of ‘determined’ and ‘determination’ here:

> Given that (as per the second moral) different sciences may treat a single determinable as having different determination dimensions, mental properties may be super-determinate [i.e. fully specified] relative to a purely psychological science, while (as per the first moral) being further determined relative to a lower-level science (treating the physical realizers of the mental properties). What is super-

> determinate relative to one science may not be super-determinate relative to another (2009: 164-165).

Upon reflection, I think it is difficult to read these and other passages, without experiencing some sense of equivocation of meaning in the uses of 'determinate of' 'determined' and such like, particularly in light of the major difficulties with trying to give an account of d/d that we have discussed above. This gives further grounds for questioning the original hypothesis that physical events might be specific ways of being mental events, and further supports the contention that d/d is not an attractive stand-alone option for R.

Finally, we return to the fourth charitable concession we gave above to the proposal that d/d–understood in terms of *specifies*–is R. We granted that *specifies* might be non-conceptual and internal, just to allow the proposal off the ground. However, we have begun to see just how charitable that was. Yablo and Wilson have failed to develop a plausible account of how *specifies* might be non-conceptual. This then invites a question: has *specifies*, as understood by Wilson and Yablo, been shown to be internal? Without the appeal to s/s, and given the failure of their arguments we have examined above, there is no way for them to give a good answer to this question. And it's certainly far from obvious that the examples of non-conceptual specification that Yablo and Wilson entertain allow a good answer either. In short, it seems obvious that *specifies* is entailed by the definitions of 'red' and 'scarlet', but it seems very far from obvious that *specifies* is entailed by the definitions of 'c-fibre activation' and 'pain', or any account of the properties, instances, or events, those terms can be used to refer to. Thus, this stand-alone account of d/d fails both the second and third explanatory criterion: it hasn't been shown to stand between the mental and the physical, or to be internal. In order for it to be shown that d/d is internal an analysis of the relevant relata is required which demonstrates that the nature of the relata entails the relation. All that Yablo and Wilson have left which might do that is the appeal to s/s, which we will discuss below.

Two further candidates for R remain to be assessed: s/s and c/c. We turn now to examine the latter.

3.6 Constituted and constituters

The two major proposals of interest that appeal to c/c each give a different account of the relation.[86] We will first consider Pereboom's account, which is the more complex, before examining Paul's.

Pereboom proposes that the mental stands in two major relations to the physical: c/c and d/d. We will examine both of these in turn starting with c/c, which Pereboom describes as the 'deepest relation between the psychological and the microphysical' (2011: 5). As we shall see, Pereboom's employment of c/c is ultimately obscurantist, and possibly circular; and his appeal to d/d inherits the problems with d/d that we discussed above. For these reasons, Pereboom's account of c/c fails the second and third explanatory criteria.

Pereboom takes his starting point from a long-standing metaphysical tradition on the nature of the c/c relation, discussion of which often focuses on the following example. Consider a statue and the lump of clay of which the statue is made. Are the lump and the statue identical to one another? Many have argued that because the lump and the statue differ with regards to their modal and temporal properties, they are not identical. The statue is destroyed by significant changes to its shape, the lump can survive such changes. What, then, is the relation that the lump and the statue stand in? Many have argued that the lump and the statue stand in the relation of material constitution, with different proponents offering different accounts of what precisely is meant by 'material constitution' (e.g: Kripke 1971; Lowe 1983, 1995; Baker 1997, 2000, 2002, 2007; Fine 2003). Pereboom's proposal is that concrete mental and physical entities stand in the relation of material constitution, and he offers a distinctive account of that relation. While discussing Pereboom, we shall take 'c/c' to refer to the relation Pereboom has in mind.

Pereboom's account of c/c has two main elements which interest us here.[87] According to Pereboom's account, x wholly constitutes y iff:

[86] For a survey of other constitution approaches, see Ney (2007); although note that Ney uses 'constitution' in a very wide sense to cover most of the views discussed here under the label 'immanentism'. The main view I don't discuss that Ney does which might best suit the label 'constitution' is Baker's (1997, 2000, 2002, 2007). Essentially, Baker's view is that constitution is a matter of necessary spatial coincidence. Perhaps the primary problem with this view is that spatial coincidence is insufficient for constitution: it seems possible for one entity to be spatially coincident with another without constituting it. On this point, Sider conceives of a world in which the laws allow two objects on a collision course to pass through one another, in such a case they could be spatially coincident yet do not constitute each other (2003: 46). In my view, Pereboom and Paul's contributions represent the state of the art for constitution approaches, hence my focus on them here.

1. x is materially coincident with y;
2. and y is wholly made up of x.

With regards to ‘wholly made up of’ (henceforth: *made of*), Pereboom says that:

> the *made up of* relation [is] basic in the sense that it cannot be fully analyzed as consisting in more fundamental relations. In particular, it has no analysis into more fundamental mereological relations. The *made up of* relation is asymmetric and irreflexive: the lattice is not made up of the diamond, and the diamond is not made up of itself. It has a specific direction: the less fundamental made up of the more fundamental. But the core of the *made up of* relation is unanalyzable and thus primitive (2011: 137-138).

Thus, other than naming some of the logical properties of the *made of* relation, Pereboom treats it as primitive. Given what we have established above, this gives us good grounds for holding that the relation is brute too. If so, then c/c can’t meet the second criterion. But let us postpone that conclusion till we have gotten a better grip on the view.

Pereboom offers two possible understandings of ‘materially coincident’. First, he considers a proposal by Dean Zimmerman that: x and y are materially coincident iff, at some level, x and y are made out of the same parts (Pereboom 2011: 138; Zimmerman 2002: 297).[88] Second, Pereboom suggests that: x and y are materially coincident iff x and y are spatially coincident. Pereboom believes that the latter is more problematic because of the arguable possibility that the material structure of an object might be distinct from the spatial structure of its location (2011: 138-139). So let us take Zimmerman’s proposal as Pereboom’s preferred understanding of material coincidence.

[87] I don’t discuss the following two aspects of Pereboom’s view. (i) Pereboom’s acceptance of asymmetric necessitation as a constraint on any appeal to c/c holding between the mental and the physical, which he acknowledges raises some questions about whether c/c can do the job. Lumps don’t necessitate statues unless they’re arranged ‘statue wise’. For reasons of space I put this issue to one side. For the record, I do think there are complications here for Pereboom’s view, but given that he appeals to the *made up of* relation, which by stipulation is a form of asymmetric necessitation, this is alone sufficient to entail dependence, which is my main concern, so it is *made up of* that I focus upon. (ii) That his position allows for the ‘dimensioned view’ of realisation defended by Carl Gillett (2002; 2003; 2007), according to which F can realise G only if F is not a property of the same object as G, but rather F is a property of a constitutive part of the object that G is a property of. This is supposed to contrast with the ‘flat view’, according to which F can realise G if F and G are a property of the same object. I hold that there is no deep metaphysical distinction between the two views.

[88] Pereboom actually says ‘just in case’ which I read as ‘iff’. Also, Pereboom denies that Zimmerman’s proposal entails that x = y, because he denies the extensionality of parthood principle, according to which if x and y are composites with the same proper parts, then x = y.

One might wonder why Pereboom appeals to both relations to explain c/c. He gives two reasons. First, that material coincidence is insufficient to explain the nature of c/c because material coincidence is reflexive and symmetric, whereas c/c is intuitively irreflexive and asymmetric. So if x is to be constituted by y, then it can't simply be the case that x is materially coincident with y; there must be more to it. Pereboom suggests that the extra ingredient is that x is made of y.

One might then wonder why Pereboom appeals to material coincidence at all, instead of just *made of*. Thus his second reason: that *made of* on its own allows for, say, the whole lump to constitute not just the whole statue but also the head of the statue–an implausible result. Material coincidence rules this out. Hence the two part analysis of c/c.

So why doesn't Pereboom's proposal work? First note that c/c entails dependence, if at all, entirely in virtue of *made of*, which Pereboom treats as largely primitive. This invites an immediate worry because although there is perhaps some intuitive sense in which the diamond is made of the lattice of carbon atoms, how are we, without any elucidation from Pereboom, to make sense of the idea that mental *properties* and *events* are made of physical properties and events like, presumably, the diamond is made of the lattice? As we will see, Pereboom struggles to provide a convincing answer to this question.

All that Pereboom says about *made of* is that it is irreflexive and asymmetric, although Pereboom must also accept that *made of* is synchronous, non-causal, modally forceful, and that the following is true of it:

A is made of B only if: necessarily, if anything x has some property F in A, then there is at least one property G in B such that x has G, and, necessarily, everything that has G has F.

If Pereboom doesn't accept these extra insights into *made of*, then the relation can't entail dependence, and thus c/c would fail the first explanatory criterion, as material coincidence can't entail dependence because it's reflexive. Somewhat puzzlingly, Pereboom does not believe that his account of c/c needs to entail dependence; as he states (in relation to another dialectic):

> it shouldn't be supposed that [c/c] would be defective if it does not guarantee some sort of [dependence] of the mental on the physical (Unpublished: 2).

However, as we have already established, if c/c does not entail dependence then it can't solve the super-overdetermination problem–which is the correct articulation of the overdetermination problem faced by NRP, a problem which is a major motivation for Pereboom's position. Unless we grant that *made of* entails dependence, Pereboom's position is immediately untenable. Moreover, how can Pereboom deny that *made of* has dependence's features?–He's already granted that it is irreflexive and asymmetric, and so surely can't deny that *made of* is synchronous, non-causal, and modally forceful. This leaves only the final condition as a possible point of contention, but once all the previous conditions have been accepted, then it's not clear on what grounds Pereboom could rule that final condition out. I cannot see how his view can be tenable at all without accepting the final condition of dependence, or something a lot like it.

So, *made of* must have the relevant properties of dependence. This invites a question: what's the difference between *made of* and dependence? Given his commitment to the primitiveness of the relation, Pereboom can give only one answer: there's an intuitive difference. Thus, whether or not there is anything informative in Pereboom's appeal to c/c rests entirely on the intuition about the 'primitive core' of *made of*, as the relation is the only part of Pereboom's appeal to c/c which entails dependence. One might be tempted simply to charge Pereboom with obscurantism regarding this, and to varying degrees some already have (e.g. Kim 2005: 61n30; Ney 2007: 504n11; Melnyk 2008: 1292-1293, 2013: 14). Pereboom denies the charge simply by saying:

> I resist the claim that this proposal amounts to obscurantism. It is *sufficiently clear* what we mean when we say that the diamond is made up of a lattice of carbon atoms, and that the brain is made up of a configuration of various kinds of neurons, even if no reductive analysis is provided for this relation (2011: 137-138; my emphasis).

Those who have charged Pereboom with obscurantism have left the charge largely undeveloped. Perhaps because the issue seems at something of an impasse: one philosopher's perspicuity being another's obscurity. As Lewis once opined: 'any competent philosopher who does not understand something will take care not to understand anything else whereby it might be explained' (1986: 203n5).[89]

But despite this appearance of an impasse among competent philosophers, there is room here for an *en passant*.[90] To see this, first note that one question that Pereboom should answer is: in response to the charge of obscurantism, you say that it is 'sufficiently clear'; but it is sufficiently clear for what, exactly? The answer, it seems, is that Pereboom thinks that *made of* is sufficiently clear to avoid the charge that the relation's nature is so opaque that no argument can rest on an appeal to it. The NRP-overdetermination dialectic has seen this type of charge before. As we saw in Chapter 2 (Section 2.6), for example, List and Menzies accused Kim of obscurantism in respect of his reliance on productionism, stating simply that: 'unless a better explication can be given of causation as production, this notion can hardly play a significant role in the debate about mental causation' (2009: 489). And Kim made a similarly brusque complaint against the original version of Pereboom's account of c/c: 'I believe that the main burden [of this approach], which is yet to be discharged […] is to produce a serviceably clear concept of constitution' (2005: 61n30).

Let us call any concept or claim which is so opaque that no argument can rest on it: 'obscure'. Applying this measure to Pereboom, it might seem at first glance that he is not being obscure. Firstly, the term 'made of' is a commonsense term whose concept, it seems plausible to claim, we all grasp to some extent. We can all have seemingly sensible enough conversations about statues being made of lumps of clay. And this appears to be part of Pereboom's point with his diamond and brain examples. To that extent, the talk of 'made of' is not obscure. But here's the *en passant*: the concern is not, or should not be, that we don't know *anything* about what it is for x to be made of y, because, of course, we at least know that *made of* shares all of dependence's properties. But what we don't know is what *made of* has *in addition* to dependence, which is what we need to know if we are to agree that *made of*, and not just dependence, holds between the mental and the physical. It is *here* that the obscurity lies.

[89] Indeed, the recent exchange between Pereboom (Unpublished: 2-3) and Melnyk (2013: 11) provides a good example of this.

[90] I refer to the chess move whereby a black pawn (say) can capture a white pawn if the white pawn moves forward two squares on its first move and ends its move directly adjacent to the black pawn; the black pawn makes the capture by moving into the square behind the white pawn, and in so doing captures the white pawn 'as it passes'.

And it is an obscurity that resists illumination. What informative things can we say about the 'primitive core' of *made of*? It is a kind of super-singularity: a supposed something about which nothing can be said. This certainly seems obscure. Thus, shorn of the properties it shares with dependence, *made of* is obscure, and there is no explanatory force in Pereboom's appeal to it.

This invites a question: what are the costs for Pereboom if *made of* just is dependence? I think there are at least two.

First cost: his view would be circular. Of course, from the perspective of the three explanatory criteria, there is, in principle, no problem with dependence being identical to *made of*. However, Pereboom does promote his view as providing an at least somewhat informative analysis of the relation between the mental and the physical, but if all that we can say about *made of* involves simply listing the already established features of dependence *made of* shares, then his view would not be informative.

All that Pereboom can do here is appeal to intuition. But such an appeal will carry little force because, as we discussed above, the dialectical weight of an intuition is proportional to its Moorean force, and the intuition that *made of* has a primitive core which distinguishes it from dependence is very far indeed from, say, the intuition that 'I *know* this is a hand'.

There are echoes here of the problem that Yablo encountered when trying to explain the nature of d/d. Yablo tried to explain dependence in terms of 'immanent' d/d, but then explained d/d in terms of dependence that was 'immanent'. Yablo at least had a potential way out of that circle, though: understanding d/d in terms of s/s. And Pereboom seems to face a similar threat of circularity: he tries to explain dependence in terms of c/c, and c/c largely in terms of *made of*–but when we examine the latter, we struggle to see how it differs from dependence. Pereboom would presumably insist that intuition tells us that dependence and *made of* are distinct; but, as we have seen, an insistence is all it is.

Second cost: if *made of* simply is dependence, then Pereboom has left us none the wiser as to whether or not *made of* and thus c/c is an internal relation–c/c can be internal only if *made of* is internal because Pereboom has offered us no account of the relevant relata which would show that material coincidence is an internal relation. Moreover, if we accept

Pereboom's claim that *made of* is primitive–that there really is the extra intuitive ingredient–then this gives us grounds for holding that *made of* is brute. Either way, Pereboom fails to meet the second explanatory criterion.

And Pereboom has also struggled to meet the third criterion because in response to the lack of intuitive plausibility in the claim that mental properties and events are constituted by physical properties and events, just like statues are constituted by lumps of clay, Pereboom has, ultimately, simply appealed to a primitive relation, *made of*, claimed its nature is intuitively graspable and left it at that. So, although we should not conclude that we have definitively shown c/c does not stand between the mental and the physical, we can at least conclude that Pereboom has not shown that we should accept that c/c does so stand.[91]

Indeed, it seems as if Pereboom has encountered a problem analogous to the one we noted that Yablo and Wilson encountered above. Recall that Yablo and Wilson struggled to give a plausible account of how the mental could stand in d/d to the physical without implicitly working with an account of d/d which treated d/d as dependence plus we know not what. Pereboom seems similarly to struggle to give a plausible account of how the mental could stand in c/c to the physical without working with an account of c/c which treats it as dependence plus we know not what.

Thus, Pereboom has ended up circling a singularity. So how did this come about? Again, we can see the effects of the NO-OD intuition at play here. As Pereboom states:

> The concern is that […] every event that is mentally caused will be overdetermined in the way that someone who is fatally hit simultaneously by each of two bullets shot by two assassins is overdetermined, a result that is implausible (2011: 28).

If one assassin constitutes the other, then one might think that the NO-OD intuition has been dealt with. But as we have seen, this is not a genuine problem for NRP; and in his efforts to 'solve' this pseudo-problem on behalf of NRP, Pereboom has tried to offer an explanation of the holding of dependence which is obscurantist, possibly circular, and fails the second and third criterion. Thus, Pereboom's view will stand or fall on his appeal to d/d. This brings us to the second part of our assessment.

[91] To be fair, Pereboom concedes that his view falls short of giving a fully physicalist account of the mental, although not for the reasons I cite here (Unpublished: 1-3).

Pereboom maintains that mental properties are 'compositional properties' which he defines as 'properties that things have solely by virtue of intrinsic features of their parts, either proper or improper, and relations these parts have to one another' (2011: 148). Compositional properties are at a higher level of abstraction than neural properties or any other properties at a lower-level. Call this 'compositionalism'. According to Pereboom, these compositional properties stand in d/d (as well as c/c) to the physical properties on which they depend. Pereboom contrasts this position with what he calls 'ER-functionalism'–short for 'external relations functionalism'.[92] According to ER-functionalism, mental properties are to be analysed in terms of their causal roles–the external causal relations that those properties stand in to other mental and physical properties. Compositionalism instead analyses mental properties as intrinsic properties.

To illustrate the general idea, Pereboom discusses an analogy with a ball piston engine, which is a kind of internal combustion engine. Ball piston engines, Pereboom notes, are characterised by possessing parts with certain shapes and rigidities, arranged in certain ways. That ball piston engines have such properties is not due to any external functional relations that the type of engine stands in. In this sense, Pereboom suggest, these properties 'comprise a compositional property intrinsic to such an engine' (2011: 153). And this compositional property is multiply realisable, Pereboom contends, because different materials could make up the relevant parts of the engine so long as the relevant shapes and rigidities were realised. Thus the analogy: the different physical realisations of, say, the belief *this fire is dangerous* that you might find in different species could 'exhibit a compositional property of a single type that is intrinsic to this kind of mental state' (2011: 154).

The basic idea would seem to be that, say, in a Martian and a human fleeing a forest fire, both would believe that the fire was dangerous in virtue of their both possessing the same compositional property which would be a property the human had in virtue of some part of his brain, and the Martian had in virtue of some part of his Martian brain (whatever that might be). According to Pereboom's model, Martian brains and human brains could differ greatly, yet, at a slightly higher level of abstraction than Martian and human brains, there might be compositional similarities between the two, in virtue of which they could both have the same mental properties. In the analogy, the Martian brain might be a ball piston

[92] 'External relations functionalism' was coined by Melynk (2006: 1292).

engine made out of steel, the human brain a ball piston engine made of titanium, yet both kinds of brain could share certain intrinsic compositional properties, just like both kinds of engine could be rigid, or shaped in certain ways.

Pereboom describes compositional properties as ‘broadly physical’, in a sense which appears similar to how I use ‘broadly physical’. One might wonder, though, whether these compositional properties are *physical*, and thus whether Pereboom is really offering a reductionist position. Pereboom denies that compositional properties are physical, although he does say that his view amounts to a ‘compromise’ with reductionism (see 2011: 148). However, I think Pereboom faces a dilemma here. Either compositional properties are physical, or they’re not. If they’re physical, then he’s a reductionist. If they’re not physical, then Pereboom is left facing the task of explaining how these non-physical properties are related to physical properties, just like every other NRPist. In which case, the extent to which this is a ‘compromise’ with reductionism is rather opaque. Pereboom will take the second horn, of course, but as we have seen his appeal to c/c *can’t* explain the relationship between the mental and the physical.

But Pereboom has not entirely run out of runway at this point, because he does say that his position can also be understood as ‘an instance’ of Yablo’s proposal that mental properties stand in d/d to physical properties:

> In the analogy, we can think of the distinctive compositional property of the ball piston engine as a determinable relative to which the steel and titanium realizations of this property are determinates. Similarly, a mental compositional property can be construed as a determinable relative to neural and silicon-based realizations of that property. Considering mental properties as sufficiently abstract compositional properties in fact makes it natural to regard the relation between them and their neural or silicon based realizations as that of determinable to determinate. Yablo’s proposal also intuitively fits with the constitution of mental property instances […] by lower-level property instances. An instance of the engine’s distinctive compositional property […] will be made up of [and] materially coincident with […] its steel-realization property instance […] and at the same time it is intuitive that the engine’s compositional property is the determinable of which the steel realization is a determinate (2011: 154-155).

This is all that Pereboom says to explain what he means by d/d. However, the little he does say at least suggests that Pereboom would accept that c/c has the all of the properties of d/d, as understood by Yablo. But in that case, Pereboom's proposal would inherit all of the problems that we have identified above with Yablo's view. Moreover, as we will see below, Pereboom cannot fall back on Yablo's appeal to s/s as that relies wholly on an appeal to a series of brute necessitation relations between distinct mental and physical properties, and Pereboom can't switch to Wilson and Shoemaker's appeal to s/s–because he explicitly rejects the claim that any mental token can be identical to a physical token (a central component of Wilson and Shoemaker's view, as we will see). As he states:

> Contemporary nonreductive physicalism about the mental is grounded in modal and temporal arguments against identity, and if one is persuaded by arguments of this sort against a type-identity thesis, one should at least be motivated to accept similar arguments against any token-identity thesis about the mental (2011: 131).

The modal and temporal arguments that Pereboom has in mind are the classic metaphysical arguments given for denying mental-physical type-identity. If it's possible to have pain in the absence of c-fibre activation, then the former can't be identical with the latter. Pereboom thinks that this reasoning transfers to tokens too: any particular instance of pain, M_1, can't be identical with its physical realiser, P_1, because it's possible to have M_1 in the absence of P_1. Moreover, Pereboom holds similar arguments will work against *any* claim to the affect that 'mental token x = physical token y'–including for token causal powers or other properties. We will consider this argument again below when discussing Paul's version of c/c as well as when we assess the s/s approach, but for now just note the consequences of this for Pereboom's appeal to d/d. As we saw above, efforts to explain d/d that did not appeal to s/s failed on a number of grounds. Pereboom has appealed to d/d, but ruled out that d/d in his sense can be understood in terms of s/s. Consequently, his view succumbs to all of the objections we considered above in relation to Yablo and Wilson, and he cannot avail himself of the fall-back position of understanding d/d in terms of s/s.

In sum, Pereboom's appeal to c/c fails because it's obscurantist, if not circular, and does not meet the second or perhaps even the third explanatory criterion. Pereboom's appeal to d/d fails because in making that move he is even more vulnerable than Yablo and Wilson are to the objections we have discussed above. Thus, we can conclude that neither c/c nor d/d, understood in Pereboom's sense, can be R.

We turn now to Paul’s account of c/c. While discussing Paul, we shall take ‘c/c’ to refer to the relation Paul has in mind.

On one reading of Paul’s view, the basic thought is that when one event stands in c/c to another event they are ‘partly identical in that they [partly] overlap by sharing many of their property instances’ (2007: 280).[93] Mental events stand in c/c to physical events in the sense that all events are bundles of property instances, and all causally relevant mental instances are identical to physical instances. Mental events are not identical to physical events, however, because they differ in respect of their non-causally relevant instances– particularly their *de re* modal instances.

For example, take the event of Alex’s migraine at noon, this will be a bundle of various mental instances, and each of the causally relevant instances will be identical to a physical instance. A migraine event might plausibly be broken down into a number of mental instances, similarly with the physical instances that make up the neural event which the mental event depends on; but, to simplify, in the case of Alex’s migraine the pain instance would be identical to the c-fibre activation instance. However, in the bundle of instances that is Alex’s migraine at noon, there would also be an instance of the modal property of being accidentally neural, whereas in the bundle of instances that is Alex’s c-fibre activation at noon, there would be an instance of the modal property of being essentially neural. Thus, although the events share the pain/c-fibre activation instance (the pain instance just is the c-fibre activation instance), the events themselves aren’t identical because they differ with regards to their modal instances.

There is a question here about how coarse-grained the events are: how many property instances are there in any event? If ‘event’ just means ‘causal relatum’, for Paul the answer is ‘one’–the causal relata simply are property instances, where ‘property instance’ stands for an object instantiating a property at a time. This immediately leads to a problem, however. If every mental event is identical with a property instance, and every mental property instance is identical with a physical property instance, then all mental events just are physical events. But if c/c is to be R, then it can’t be the case that all mental causal relata–i.e. events–are identical to physical causal relata, because, even though identity may

[93] Paul takes property instances to be the causal relata, I continue here to treat events as the causal relata; nothing hangs on this.

indeed hold between the mental and the physical (in my view), identity does not entail dependence. So c/c could not be R because it would fail the first explanatory criterion.

One might think that Paul's view has the resources to deal with this issue, however, in virtue of the fact that not every mental instance is identical to a physical instance – at least modal mental instances are non-identical. Consider Paul's example where she asks us to consider an occurrence of protein 'Pro' and an occurrence of the sum of molecules that constitute it 'Mol':

> The nonreductionist [could] hold that the property instance of *being protein Pro* is a complex property instance that is really just a conjunction of many more fundamental property instances such as *having shape s*, *having mass m*, *having n molecules accidentally*, and so on. Now consider the property instance of *being sum Mol*: it is a complex property instance that is just a conjunction of many more fundamental property instances such as *having shape s*, *having mass m*, *having n molecules essentially*, and so on. Just as *Pro* and *Mol* partly overlap or share their property instances, the property instance of *being Pro* and the property instance of *being Mol* share some of their conjuncts […] Further, as a conjunctive property instance such as *being Pro* is just the fusion of the conjuncts, likewise the property instance for *being Mol* (2007: 284).

Applying this to Alex's migraine, the property instance of being Alex's migraine at noon would be a complex property instance which is a conjunction of many more fundamental property instances, F, G, H, which are also conjuncts of the property instance of being Alex's c-fibre activation at noon, yet the two instances would differ with regards to their respective modal conjuncts. Thus, on this understanding of Paul's view, pain and c-fibre activation instances, and thus events, wouldn't be identical, however it's not immediately clear how this might enable the required asymmetry such that c/c might entail dependence: the *overlapping* relation is symmetric. What I take Paul to have in mind, though, is that each mental event, ME, overlaps a range of physical events, PE_1, PE_2… PE_n, such that, even though ME stands in a symmetrical overlapping relation with each of PE_1, PE_2… PE_n, the overall relation between ME and the PE_1, PE_2… PE_n is asymmetric: if ME, then no particular event from the range of PE_1, PE_2… PE_n is necessitated, but if PE_1, PE_2… PE_n, then a particular event, ME, is always necessitated; and therein lies the asymmetry.

This raises a major issue that we will postpone further discussion of until Chapter 4 (Section 4.1). Namely that Paul's view is a disjunctive physicalist view, about events at least (Paul officially endorses NRP, but does not try to explain the relationship between mental and physical properties, *per se*). There would seem to be identity licencing bridge laws between ME and the disjunction of PE_1, or PE_2… or PE_n of the form: ME iff PE_1, or PE_2… or PE_n. Moreover, according to Paul, the only difference between ME and the PEs is the fact that ME is not essentially any one of the PEs. Thus, Paul's view allows for a disjunctive physicalism in respect of events at least.

Just now, we will focus on another significant problem faced by Paul's view: it relies on a whole host of epistemically brute, trans-ordinal identity claims. Recall that a brute identity claim, in epistemic terms, is a claim of the form 'A=B' where no true, informative explanation can be given for why we should believe that A=B. We have discussed above how reductionists cannot simply make epistemically brute identity claims, but here we find a non-reductionist doing just that. Notice that Paul simply helps herself to the claim that, in the case of the complex property instance of being Alex's migraine at noon, all of the causally relevant mental conjuncts of that property are identical to a physical conjunct of the complex physical property instance of being Alex's c-fibre activation at noon. But why should we accept this? Consider one of the mental property instances which are one of the conjuncts of Alex's migraine: being searing to the maximum degree. Which physical instance which is a conjunct of Alex's c-fibre activation is the searing instance identical with, exactly? Paul provides no answer to questions like this, and there are no resources to answer them in her view. This is because Paul has simply made an epistemically brute identity claim: we are told that the searing instance *just is* identical with some physical instance.

And Paul is in a particularly difficult position because if non-reductionists can help themselves to trans-ordinal brute identity claims like this, then why can't reductionists? What's good for the goose is good for the gander. But non-reductionists are supposed to be motivated in large part by a *rejection* of implausible mental-physical identity claims. So a non-reductionist view which is based on implausible mental-physical identity claims does not seem to be much of an improvement. Moreover, in making a brute identity claim, Paul has failed to meet the second and third criteria: brute identities can't be shown to be internal.

Also, once we see that Paul has simply made a brute identity claim, she becomes particularly vulnerable to the kind of multiple realisability argument against mental-physical token identity that Pereboom put forward above. Take the instance of being searing to the maximum degree, M_1, and the physical instance with which it is supposed to be identical P_1. If $M_1 = P_1$ then it's impossible to have M_1 in the absence of P_1–but surely, if NRP is true, that *is* possible. If, say, P_1 were instantaneously teleported out of Alex's brain without damage, and instantaneously replaced with a distinct P_2 that functioned in the same way, then it's not clear why M_1 would cease to exist–nothing *mental* would have changed. And that's the point of NRP: the mental can be multiply realised by the physical. As Pereboom points out, one cannot, in a principled manner, restrict multiple realisation to types only.

Finally, I think that the NO-OD intuition's pernicious effects can be seen in the problems that Paul faced, in a way similar to what we have seen above. Consider when Paul states that:

> Whether we focus on objects or properties, once we realize that it is the involvement of certain property instances that determine whether one thing causes another, and that in the cases of constitutive overdetermination that we have considered, the causally important or relevant property instances are shared, we can see why causal responsibility is shared, not overdetermined. When I, my constitutive sum of cells, my constitutive sum of molecules, and so on, hit the tennis ball, a *shared part* (a shared property instance) causes the tennis ball to bounce of my racket at 100 miles per hour (2007: 285).

What Paul is constructing her metaphysic to rule out is that there might be *two* causes in any problematic sense; on that front Paul may well have succeeded. However, as we have seen, this is a solution for a non-problem. And although Paul's view, along with the other immanentists, at least promised an explanation of the dependence relation between mental and physical causes, we see here that promise is unfulfilled.

Only one candidate for R remains: s/s, and we now turn towards an examination of it.

3.7 Sets and subsets

The two major proposals that appeal to s/s each give a different account of the relata. We will first consider Yablo's account, before examining Wilson and Shoemaker's. While discussing Yablo, we shall take 's/s' to refer to the relation Yablo has in mind.

As we will see, Yablo's explanation of the necessitation relation between mental and physical events relies wholly on an appeal to a series of brute necessitation relations between distinct mental and physical properties. Thus, Yablo's appeal to s/s is a dead-end: it does not provide a way out of the circle that we identified above, and does not provide any further resources for making sense of the claim that physical events specify mental events. Moreover, by appealing to brute necessitation between mental and physical properties, Yablo's view fails the second and third criterion, and neither s/s, nor d/d understood in terms of s/s, can be R.

Yablo contends that the essences of mental events are non-empty proper subsets (henceforth: subset) of the essences of the physical events on which those mental events depend. The properties without which an event, E, cannot exist are the essential properties of E, and E's essence is a selection of those essential properties. The selection includes the properties required to make E the event that it is, with the exception of properties like being self-identical, and being of type T, which, according to Yablo, are necessitated by E's essence, but are not part of it. Yablo wants to exclude such properties because they would frustrate the attempt to explain the non-reductive relationship between mental and physical events in terms of s/s. For example, the essence of a pain event couldn't be a subset of the essence of a c-fibre activation event if the former's essence included *being identical to a pain* event. The only properties that can be included in event essences, according to Yablo, are categorical properties. The general idea is that a property is categorical iff its possession by an event at a world is 'strictly a matter' of that event's nature at that world, not a matter of how the event would or could have been (1992: 261-262).[94] Non-categorical, hypothetical properties include counterfactual and modal properties.

This gives us the following model of the relationship between, say, Alex's noon migraine and the c-fibre activation on which it depends:

[94] There is a circularity issue here that I do not address for reasons of space (see Yablo 1992: 263-264).

Event	*Essences*
Alex's migraine at *t*	$\{F_m, G_m\}$
Alex's c-fibre activation at *t*	$\{F_m, G_m, H_p, I_p, \ldots\}$

Here, the properties F_m and G_m are mental, and H_p and I_p are physical. According to Yablo, the physical properties of the physical event (H_p and I_p in the above case) necessitate the mental properties (F_m and G_m in the above case). As he states:

> Every mental property *Mk* in *m*'s essence is backed by a necessitating physical property *Pk*; and as before, these physical properties attach also to some realizing physical event (this time called *q*). Even if some or all of the *Pks* are only accidental to *q*, we can imagine a more determinate physical event *p* to which they are all essential. But then *p* has essential physical properties to necessitate every mental property in *m*'s essence; and it follows that these mental properties are in *p*'s essence too (1992: 267-268).

Thus, according to Yablo, a physical event necessitates a mental event entirely in virtue of the necessitation relations between the physical properties in the physical event's essence and the mental properties in the mental events essence. The problem is that this puts us right back where we started: with a necessitation relation between mental and physical properties that we need to explain the holding of. However, Yablo does not give such an explanation, and, ultimately, provides no resources in his view that might allow one to give an explanation on his behalf. Thus, the necessitation relations that Yablo appeals to are at least epistemically brute, and so his appeal to s/s, or d/d understood in terms of s/s, fails the second and third criteria.

Moreover, it is far from clear how this appeal to s/s justifies the claim that physical events specify mental events. Asymmetric necessitation, as we have seen, is not sufficient for specification. But, ultimately, the appeal to s/s boils down to an appeal to patterns of asymmetric necessitation relations between the relevant properties of the respective essences, which in turn generate asymmetric necessitation between mental and physical events, so the claim that mental events are specified by physical events remains unjustified.

Finally, we might summarise Yablo's overall view as follows. The mental depends on the physical. That dependence relation is to be explained in terms of 'immanent' asymmetric necessitation. Immanent asymmetric necessitation can be explained in terms of a non-conceptual variety of d/d. That non-conceptual variety of d/d can be explained in terms of asymmetric necessitation that is 'immanent'. There is also a relation, s/s, which holds between the essences of mental and physical events, but not between mental and physical properties, although it is analogous to the d/d relation that holds between mental and physical properties. The holding of the s/s relation between mental and physical events is explained by asymmetric necessitation relations, or, as I like to say, dependence relations that hold between mental and physical properties. Thus, the circle is complete.

We turn now to consider the account of s/s given by Wilson and Shoemaker. I argue as follows. First, Wilson and Shoemaker fail the second criterion because they appeal to brute identities between mental and physical causal powers–identity relations upon which their entire view depends. Second, the view fails to give a convincing account of d/d in terms of s/s, and so, as we saw with Yablo, this confirms the failure of the stand-alone d/d version of immanentism. Third, Shoemaker's view is best understood as a version of reductionism, rather than non-reductionism.

While discussing Wilson and Shoemaker, we shall take 's/s' to refer to the relation Wilson and Shoemaker have in mind.

There are a number of details of and differences between Wilson and Shoemaker's account that will not be discussed here (see Wilson: 1999; 2009; 2011; Shoemaker 2001; 2003; 2007; 2013). The key claim that Wilson and Shoemaker agree on is that the causal powers associated with any mental property are a subset of the causal powers associated with each of the physical properties that the mental property depends on, or is realised by.[95] The causal powers associated with each property are called the 'causal profile'–and the subset relation holds in virtue of relations between the causal powers in the causal profiles of the relevant properties. On the most straightforward version of the view, that subset relation holds in virtue of every causal power of the realised property being identical with a causal power of the realiser property. As we will see below, Wilson and Shoemaker put the point

[95] There are different ways of reading how this claim might be understood in relation to events, but for present purposes we need only get to grips with the claim as it stands. Also, note that Shoemaker uses the term 'causal feature' instead of 'causal power' – I use the latter term for ease of exposition only; nothing will hang on it.

in terms of the identity of causal power tokens. I start with the most straightforward version of the view to enable a clear statement of the general approach, and to allow comparison with the more nuanced versions of Wilson and Shoemaker.

The subset model is supposed to offer an analysis of realisation, as well as d/d.[96] The analysis suggests the following model of the relationship between a mental property and its physical realiser properties:

Property	*Causal Powers*
Mental Realised	{CP1, CP2}
Physical Realiser 1	{CP1, CP2, CP3, CP4}
Physical Realiser 2	{CP1, CP2, CP5, CP6}
Physical Realiser 3	{CP1, CP2, CP7, CP8}

One might ask at this point: what is a causal power? This is a difficult question to give a complete answer to, but the basic idea is that a causal power of a property is itself a kind of property–a property of a property. Some will no doubt be uncomfortable with talk of properties having properties, but Shoemaker describes it as ‘obvious’ while noting that causal powers are ‘properties of a different sort than those they characterize’ (2007: 12n4); Wilson is officially neutral on the question of whether or not causal powers are second-order properties (2011: 127), but often talks as if that were the case (2011: *passim*), so let us illustrate the idea of a causal power along these lines. Causal powers are ‘conferred on’ or ‘bestowed upon’ objects or events by the properties that possess those powers, in something like the following way. Consider the property of being knife-shaped (this is based on Shoemaker’s own example, 2007: 24). The property of being knife-shaped confers on any object that has it the causal power to cut things. This is ‘conditional’–only when the property of being knife-shaped is combined with other properties is there a causal power bestowed *simpliciter*. So being knife-shaped confers the power to cut things only if the object has other relevant properties. For example, if the object has the property of being made of silk, being knife shaped won’t confer the power to cut very much, if anything; whereas if the object has the property of being made of titanium, then the power to cut many things is bestowed (granting still further properties like being of sufficient size and so forth).

[96] Here, treat all talk of ‘realisation’ is interchangeable with talk of ‘dependence’ – arguably, the talk might not be interchangeable, but nothing will hang on it.

As noted, this is the more straightforward version of the s/s view, in particular I have suppressed mention of property instances–I consider the view cast in those terms below. The current articulation in terms of properties will suffice for present purposes. With this understanding of causal powers in play, we might now ask: what's the relationship between a property and its causal powers? Shoemaker is sympathetic to two positions: that every causally efficacious property is either identical to its causal powers (1980), or has its causal powers essentially (1998), although both he and Wilson contend that the s/s view does not require buying into either position, but does require at least holding that properties have their causal powers with nomic necessity (Shoemaker 2007: 142; Wilson 2011: 132-133). This is what the talk of a property and the causal powers it is 'associated with' is supposed to gloss over: whether the property just is the relevant bundle of causal powers, has those powers essentially, or the weaker nomic claim. Let's consider the three options.

First: every mental causal power is identical to a physical causal power. But note: these are trans-ordinal brute necessities–in this case identities. The claim is being made that A=B, but it does not appear that A=B, and we are given no true, informative explanation for why we should believe the identity–it's simply an assumption of this version of the s/s view. And, epistemically, the As are mental, the Bs are physical, so the '=' is epistemically trans-ordinal. Take the property of being in pain, this has a number of plausibly mental causal powers that it might confer upon any object or event that instantiated the property of being in pain (assuming the relevant additional conditions are satisfied). For example, the power of causing a desire not to be pain. But it is far from obvious which physical causal power this could be identical with. This is because it's not normally obvious, if it ever is, if the truth has been spoken when claims are made about this or that mental entity being identical to this or that physical entity. If such matters were obvious, then the mind-body problem wouldn't be much of a problem. Moreover, given NRP, it is even less obvious which physical causal powers had by each and every physical property which realises pain are identical to the respective causal powers of pain. To paraphrase the Putnam quote we encountered in Chapter 1 (Section 1.3):

> If we can find even one psychological causal power predicate which can clearly be applied to both a mammal and an octopus (say 'the power to cause eating'), but whose physical-chemical 'identical' causal power is different in the two cases, the view that mental causal powers are identical to physical causal powers has

> collapsed. It seems to me overwhelmingly probable that we can do this. (For the original quote, see Putnam 1967:165).

It's worth paraphrasing Putnam here because it's a classic example of how NRP began with a general concern about the possibility of identifying mental and physical entities. It is interesting, therefore, to see how far NRP has travelled from that starting point. We will return to this thought below.

The main motivation that Wilson and Shoemaker have for putting forward the subset view is that it defeats the NO-OD intuition. We don't need to rehearse again how misplaced that motivation is. To be fair, though, we should note the other main arguments given by Wilson and Shoemaker for why we should accept mental-physical causal power identity: that it's sufficient for physicalism (Wilson 2011: 129), and that it's necessary for physicalism (Shoemaker 2007: 9). And I take no issue with such arguments here, *per se*, but what I do contend is that to be a non-reductive physicalist, one must meet the three explanatory criteria. But s/s cannot be R if its holding relies on brute trans-ordinal identity claims. This trumps any other considerations, especially those motivated in response to the NO-OD intuition.

An important point to note here is that on this model mental and physical properties have largely been reduced to mental and physical causal power properties. So any questions we may have about why we should believe that any relation R holds between the mental and the physical is really, on this model, a question about why R holds between mental and physical causal power properties. Consequently, apart from the framing in terms of causal power properties and how that affects patterns of asymmetry between the different ontological levels, it is difficult to discern any significant difference between this version of the subset view and old-school brute-identity claim reductionism of the sort we saw Christopher Hill criticise at the beginning of this chapter. The crucial point is that both views rely on epistemically brute trans-ordinal identity claims, and to that extent are equally objectionable.

But what if properties are not identical to their causal powers? This brings us to the second and third options: hold that mental and physical properties have their causal powers with metaphysical or nomic necessity. However, this does little to help the subset view meet the three criteria: we still have mental causal powers which are identical to physical causal powers. The epistemically brute identity remains.

One might respond by pointing out that the s/s relation holds between causal power tokens, not types. In other words, it's the instances of the causal powers of a mental property that are identical with the instances of the causal powers of the physical realisers, not the causal power properties themselves. Wilson and Shoemaker do both state that s/s holds between causal power tokens, not types (Wilson 2011: 126; Shoemaker 2007: 19-20). But this does not allow a response against the above objections: the bruteness is the issue, not the type/token distinction.

And we can also see here how vulnerable Wilson and Shoemaker are to Pereboom's objection that any mental token can be multiply realised too. If mental causal power tokens can be multiply realised, then they can't be identical to physical causal power tokens. Wilson tries to respond to this by claiming that it's as absurd to claim that a mental causal power token could be multiply realised as it would be to claim that a fully determinate shade of red could be realised by another shade of red (2011: 140). I find this response as puzzling as it is question-begging.[97]

Consider, other mental tokens can be multiply realised: an instance of the property of being Alex's migraine at noon could be realised by a neural instance, or a silicon instance or some other instance (imagine instantaneously swapping parts of Alex's brain with non-neural but functionally isomorphic silicon implants). As Pereboom notes, such arguments are a core motivation for NRP. So why not causal power instances? Simply comparing them to determinate shades of red does not provide a persuasive answer.

Thus, Wilson and Shoemaker's view fails to meet the second and third criteria in a way similar either to Yablo's or Paul's: brute necessitation relations between mental and physical properties, or instances. Like Yablo and Paul, Wilson and Shoemaker do not give an explanation for why we should accept that these identities hold, and provide no

[97] Even more puzzling is Shoemaker's apparently unprompted admission (not in response to questions about multiple realisability), that mental causal powers can be multiply realised by physical causal powers. This is puzzling because Shoemaker is appealing to causal powers to explain the realisation relation, so to posit that realisation holding between causal powers (which he acknowledges do not themselves have causal powers) is a curious move. Shoemaker suggests that it is realisation in another sense, but does not elucidate what this second sense of 'realisation' is (see 2007: 20, 20n4).

resources within their views that might allow us to give an explanation on their behalf. So their appeal to s/s, or d/d understood in terms of s/s, fails the second and third criteria.

Moreover, in an echo of our critique of Yablo, it is not obvious how this appeal to s/s justifies the claim that physical properties (or events) specify mental properties (or events). Asymmetric necessitation, as we have seen, is not sufficient for specification. Ultimately, Wilson and Shoemaker's appeal to s/s boils down to an appeal to a pattern of identities between mental and physical causal power properties, or instances, which in turn generate asymmetric necessitation relations between the relevant mental and physical properties, or events. But this lends little weight to the claim that mental properties, or events, are specified by physical properties, or events.

Finally, let us consider the question of the extent to which the subset view is non-reductive. There are two immediate directions in which one can head if one were to abandon NRP, and Wilson and Shoemaker express sympathises in each of those directions: towards emergentism, and towards reductionism, respectively. This, in microcosm, is the issue NRP has faced from the beginning – making good on its promise of a 'third way', without slipping back onto the reductionist or emergentist path.

Consider, Wilson has recently declared:

> I am neutral on whether NRPism is the correct view about mental and physical states, since I think robust emergentism remains a live empirical possibility […] Of the NRPist options, however, I think the determinable/determinate relation is the most promising (2011: 150n10).

And Shoemaker has recently come close to declaring that his version of the subset view is consistent with the possibility that mental properties are identical to physical properties; as he states:

> Associated with any mental property there will be one physical property that *seems* an especially good candidate for being identical with it […] Every property instance has a microphysical realizer of a type whose causal profile exactly matches that of the property of which it is an instance, and the thing having that property instance thereby has the [microphysical] property [what Shoemaker calls an MSE*

> property]. These MSE* properties should count as physical properties, given that their instantiation just consists in microphysical entities being propertied and related in certain ways […] it is worth asking what the implications would be for the philosophy of mind if it should turn out that […] all properties of persisting things, including mental properties, are MSE* properties, and so are physical properties […] Would this undermine the case for non-reductive physicalism? It would do so […] if it's made definitive of non-reductive physicalism that it denies any sort of type identity between mental properties and physical ones. But I don't think this should be made definitive of non-reductive physicalism [because accepting] such an identification would not in the least support the claim that psychology is reducible to physics, and would not put into question the autonomy of psychological explanation [thus, NRPists] should be happy to accept psychophysical property identities involving MSE* properties (2013: 53-54; my emphasis).

Note here that Shoemaker might seem to have aligned himself with disjunctive physicalists like Clapp and Antony that we concluded were reductionists in Chapter 1. Indeed, Shoemaker explicitly states that his view is 'similar' to Clapp's (2007: 11). However, Shoemaker does start the above quote with the claim that 'there will be one physical property that *seems* an especially good candidate for being identical with it' and then follows it by saying that, due to the possibility of dualist worlds, mental properties can't be identical to physical properties, and so can only be realised by them (2013: 55).

Although it should also be noted that Shoemaker does endorse the claim defended by disjunctive physicalists that mental-physical property identity is not sufficient for reduction. With regard to this claim of Shoemaker's, we face a choice: either we agree with Shoemaker, Clapp and Anthony that people like J. J. C. Smart and Herbert Feigl are not reductionists, or we should disagree. As we have established in Chapter 1, we should do the latter.

Consequently, the s/s view offered by Wilson and Shoemaker is a reductionist position of the disjunctive physicalist sort. This is because, for any mental property M, and the realisers of M which have as a subset the causal powers of M, a bridge law of the following form would hold 'M iff P_1,or P_2… or P_n'. But if such bridge laws do hold, then an identification of M with the disjunction of its realisers is licensed, and disjunctive physicalism, and thus reduction, follows.

3.8 Conclusion

What lesson are we to draw from all of this? I think we should recognise the extent to which reductive and non-reductive physicalists face a similar challenge: avoiding trans-ordinal bruteness. This is the issue at the heart of the super-overdetermination problem. However, reductionists face an epistemic version of this problem, non-reductionists a metaphysical version (in addition to the epistemic one). To that extent, non-reductionists face the greater difficulty. Indeed, we have seen just how much immanentist struggle to deal with that difficulty in this chapter, and that, Pereboom's obscure view excepted, they have all ended up making brute necessitation claims of one sort or another. But a metaphysically brute trans-ordinal necessitation claim is worse than an epistemic one, the former is directly inconsistent with NRP, the latter is not directly inconsistent with RP (although it may be unacceptable for other reasons). We also saw how most of these immanentist views were versions of disjunctive physicalism. This invites questions about what approach is best for physicalists to take, with reductionism being the obvious alternative to NRP. We examine the prospects for RP versus NRP in our fourth, and final chapter.

Chapter 4
Physicalism entails Reduction

Chapter Summary

In the last chapter, we considered the cutting edge of NRP: immanentism, which promised a solution to the super-overdetermination problem at least in so far as immanentists offered explanations of the relationship between mental and physical causes. However, we saw that no version of immanentism met the three explanatory criteria that were laid down, with all failing to meet at least the second criterion: that R be internal.

In this, our fourth and final chapter, we recap the NRP versus RP debate in order to get a picture of the state of play. Namely, that the best version of NRP is classic non-reductionism, and the best version of RP combines the mental-physical relata analysis offered by realiser functionalism with the multiple realisation and autonomy accommodating features of disjunctive physicalism. Given that there is not a great deal to choose between these two variants of RP and NRP (I eschew what would seem to be a rather blunt appeal to parsimony on behalf of RP), I then put forward two general problems for NRP, to show why RP should be the preferred view for physicalists: the instance problem and the ghost problem. The instance problem is that mental instances must be identical to, proper parts of, or have as proper parts physical instances; but each option entails that mental instances must be physical. But if mental instances are physical, then mental properties are physical properties. The ghost problem is that mental instances are extended, but physicalism is inconsistent with extended non-physical entities – ‘ghosts’. So mental instances must be physical. If mental instances are physical, then mental properties are physical. Thus, I conclude, if we are to be physicalists, then we need to be reductive, rather than non-reductive physicalists.

I shall often speak of it, with deliberate abusiveness, as 'the dogma of the Ghost in the Machine.'
- Gilbert Ryle[98]

4.1 Reduction versus non-reduction: the state of play

I think that we have seen enough in the last three chapters to motivate a defence of reductive physicalism. Let us retrace some key steps to bring the point fully into view. We began by considering what mental-physical reductionism is: a mental-physical property identity thesis. This contrasted with non-reductionism, the thesis that mental properties are not physical properties. One view in particular caught our interest: disjunctive physicalism, which we deemed a reductionist view, despite what its proponents contended. We then considered how to correctly articulate the overdetermination problem that NRP is supposed to face, before considering immanentism, which seemed to offer the possibility of a solution to the correct articulation of that problem on behalf of NRP. But as we saw, none of the immanentist proposals could meet the three explanatory criteria, and many struggled to maintain their non-reductive credentials in the process. Let us take a moment to explore that latter issue.

Yablo and Wilson's appeal to d/d was beset with problems, the discussion of one of which, recall, we postponed until this chapter (see Section 3.4). That problem was that the d/d approach, if it worked, would licence a move towards reductive physicalism in the manner positively argued for by disjunctive physicalists: the disjunctive predicate the disjuncts of which picks out the determinates of a given determinable can plausibly be treated as a properly disjunctive predicate and as naming a natural kind. Thus, if mental properties are determinables of their physical realisers, then the disjunctive predicate which picks out the physical determinates said to realise the mental property can name a natural kind, and the mental property can be identified with that disjunctive, physical, natural property.

Paul's appeal to c/c faced a similar worry (see Section 3.6): she explicitly embraced commitments that lead to disjunctive physicalism about events (whilst remaining largely silent on mental and physical properties). On Paul's view, any type of mental event, ME, differs from the physical events that realise it, PE_1,or PE_2… or PE_n, only in so far as ME

[98] (1949: 15-16).

could have been realised by any one of the physical realisers; and apart from that difference, ME overlaps (i.e. shares all the same property instances as parts) with each of PE_1,or PE_2… or PE_n. But in that case, Paul’s view licences bridge laws between ME and the disjunction of PE_1, or PE_2… or PE_n of the form: ‘ME iff PE_1,or PE_2… or PE_n’, and there is no obvious obstacle to the disjunction of PE_1,or PE_2… or PE_n being considered a natural kind. It might be objected that the events picked out by each disjunct of the disjunction would be wildly heterogeneous *vis-à-vis* one another, but on Paul’s view this is not the case because of the massive degree of overlapping between ME and each of its realisers. If the realisers all massively overlap with ME, then they all massively overlap with each other. This invites the disjunctivist move: the mental predicate which designates ME is coextensive with the disjunctive physical predicate which designates PE_1,or PE_2… or PE_n, thus the bridge law holds and the identification of mental events with a disjunction of their realisers is licensed. Moreover, given Paul’s silence on the relationship between mental and physical properties, she says nothing to rule out an analogous disjunctivists move in that latter respect either.

And as we have just seen at the end of Chapter 3 (see Section 3.7), Wilson and Shoemaker’s appeal to s/s, in identifying mental causal power properties with physical causal power properties, similarly invites reduction by building into the axioms of the view that mental causal power properties are identical to physical causal power properties, in a way that, if it accommodates multiple realisation, does so in a disjunctive physicalist manner. Consider: the bridge law ‘M iff P_1,or P_2… or P_n’ would hold between the mental property M, which is, or is associated with the set of causal powers {C1, C2, C3}, and the realisers which are, or are associated with, the relevant set of physical causal powers for each realiser–all of which, by hypothesis, include {C1, C2, C3}. And once that is accepted, disjunctive physicalism is up and running via the standard appeal to the bridge law.

One potential response here on behalf of Shoemaker and Wilson would be to contend that mental properties entail, but are not identical with the relevant causal powers, and so have further quiddities which allow them to be distinguished from physical properties, thus blocking reduction. Putting the mystery mongering nature of this move to one side, it’s clearly too strong: insisting on mental quiddities in this way would also work to save non-physicalist views from physicalist anti-dualist objections. Consider, for example, how an emergentist could insist that there must be brute laws of emergence because of the quiddities of mental properties which, by dint of the mysterious nature of these quiddities, can only be linked to the physical via brute laws of emergence which we must simply accept with natural piety. Thus, physicalists employ an argument from mental quiddities at their peril.

The two most clearly non-reductionist positions that we considered were Pereboom's appeal to c/c and Yablo's appeal to s/s. The former, we saw, was obscurantist, and, confusingly, explicitly identified itself as a kind of d/d view. If it is a d/d view, then it would also licence the disjunctive move, in the way we began this section by discussing; if it is not, then it remains marooned in obscurity. Yablo's appeal to s/s included the explicit commitment of mental-physical property non-identity, but only at the price of brute necessitation relations between those properties: this left Yablo right back where he started, and us none the wiser.

Indeed, we also saw how immanentists relied on brute necessitation claims. This, we concluded Chapter 3 by observing, invites questions about the viability of NRP in comparison to RP. This is particularly true given that immanentists, who constitute NRP's cutting edge, have been revealed to be a rather motley crew–defending either some variant of disjunctive physicalism, a position whose non-reductive credentials are hard to defend, or a more robust non-reductionism which can be rejected. Thus, we have reached the summit, and are now in a position to assess the overall state of play in the NRP vs. RP debate.

I think that state of play is as follows. For our purposes, there are three varieties of NRP.

First, there is the variety which does justice to its non-reductionist label by including robust commitments to mental-physical property non-identity, but simply fails on its own terms, in particular by appealing to brute trans-ordinal relations. In this category fall Yablo's appeal to s/s and Pereboom's appeal to c/c.

Second, there is the variety which does not do justice to its non-reductionist label because it includes explicit commitments which directly lead to disjunctive physicalism, or fails to rule the option out. This variety also relies on brute trans-ordinal relations. Into this category fall Yablo and Wilson's appeal to d/d, Paul's appeal to c/c, and Shoemaker and Wilson's appeal to s/s.

Third, there is classic NRP, which includes a robust commitment to mental-physical property non-identity, and also provides an internal relation, namely causal role-playing, and so avoids appealing to brute trans-ordinals. Recall that if a mental property, M, is defined in terms of its causal role, and we empirically discover that a range of physical properties can play that role, then the relation *plays the causal role of* is entailed by the nature of the mental property and the relevant physical properties.

In light of the considerations we have put forward above, classic NRP looks like the best bet for non-reductionists. Given our conclusions about how the problem of overdetermination, traditionally conceived, is not a genuine problem for NRP, it should perhaps not surprise us that classic NRP is non-reductionism's stand-out candidate. After all, it was concerns about overdetermination that formed one of the main motivating factors to move away from classic NRP towards immanentist views – the latter of which bet the farm, and the townhouse too, on defeating the NO-OD intuition. Given what we have established above, this was a mistake, and marks a significantly erroneous turn in the history of NRP.

In contrast to this varied picture for NRP, we have reductionism, the main problem for which, as we have seen, is not multiple realisability, or mental autonomy, but rather avoiding epistemically brute identity claims. In Chapter 1, we considered one version of RP which seemed to do this: realiser functionalism. This suggests a way forward for RP: adopt the kind of analysis of mental-physical relata realiser functionalists employ, to avoid epistemic bruteness, and combine it with a disjunctive physicalist metaphysic to accommodate multiple realisability and mental autonomy. In short, hold that mental properties are functional properties which are identical with a disjunction of the physical realisers which perform the relevant function. This looks like RP's best bet.

So this is the position we have reached: classic functionalist NRP versus realiser functionalist, disjunctive RP.[99] There might not seem to be a lot to choose between these

[99] All of this, of course, leaves to one side the rather difficult question of whether or not functionalist analyses of mental properties are any good, for some of those properties or all. The qualitative aspects of mental states, the painfulness of pain, and so on, are widely regarded to resist such an analysis. I do not address that debate here as I see it as a problem for functionalist physicalism across the board, rather than a problem for NRP or RP – and it is the dispute between NRP and RP that I am adjudicating. My contention is that physicalists need to avoid brute trans-ordinals, and that functionalist analyses seem to be the best way of achieving that, currently. This leaves open, of course, the possibility that non-functionalist (or quasi-functionalist) analyses might be offered which allow an alternative way forward.

two views, but we will now consider two problems faced by NRP, which support the contention that RP is to be preferred: the instance problem, and the ghost problem.

4.1 The instance problem

My argument in brief: NRPists should accept that mental property instances are identical to, parts of, or have as proper parts physical property instances. But to accept that is to accept that mental properties are physical properties. This supports the argument that NRP is an unstable position – it must topple either into non-physicalism, or reductionism.

Consider: properties have instances. Properties like *being in pain* or *being a colour* can be found here, there, and everywhere–or, at least, their instances can. Indeed, the closest anyone ever gets to a property is to come across an instance of one: Alex's migraine at noon, Alex's c-fibre activation at noon, the redness of Alex's temple at noon, and so on. Thus, a question: what is a property instance? The intuitive notion of a property instance is best illustrated with examples: given this or that property, an instance of that property is something we can point to – here's an instance of being the letter 'z'; here's an instance of blackness: •; and so on. Intuitively, then, instances are singular entities: talk of '*a* property *instance*' suggests that instances are particulars.

For those who are not satisfied with the above intuitive characterisation, we can say something more; namely that a property instance is an object, x, instantiating a property, F, at a time, t: [x, F, t]. This is a commonly used account of property instances, according to which such instances are also events (Kim 1966). In what follows, I do not endorse the claim that an object instantiating a property at a time is an event, although I pick no fight with those who do wish to endorse such a claim. My target is different: arguing from instance identity to property identity; thus I lay to one side the issue of whether instances in this sense are events. The key point is that the entities of interest here are property *instances* (whether or not those instances are events).

The dominant version of classic NRP holds that mental tokens are identical to physical tokens. This is the token identity theory. If instances are tokens, and they would seem to be, then token identity theorists are also instance identity theorists (although they may also hold that other kinds of mental and physical tokens are identical too, such as events).

Consider a pain instance: Alex's migraine at noon; this pain instance, according to the token identity theory, is identical to a physical instance, say Alex's c-fibre activation at noon. In what follows, identity theory will be a major target, but I will also aim at two other views which might be retreated to if token identity is shown to be untenable for an NRPist. Namely, the view that mental instances are parts of physical instances, or that mental instances have only physical instances as proper parts. As I use the term 'proper part', x is a proper part of y iff x is a part of y and x ≠ y.

Thus, there are three possibilities for the relationship between mental and physical instances that are of interest here: (i) mental instances are identical to physical instances; (ii) mental instances are proper parts of physical instances; (iii) mental instances have only physical instances as proper parts. Token identity is a widely held view among NRPists. Pereboom's position might be read as a version of the view that mental instances have physical instances as proper parts; as he puts it – the mental is made up of the physical. The claim that physical instances have mental instances as proper parts is simply an alternative it seems worth ruling out – if token identity is shown to be problematic for NRP, and mental instances not to be made up of physical instances, then an obvious immediate rejoinder is to claim that mental instances are parts of physical instances.

The problem for NRP is that all of these options entail reductionism. First, let's consider the identity option, starting with a question – what are the identity conditions for instances? If mental instances are to be identical to physical instances, then we need to know what conditions need to be met for that to be. A popular proposal in that regard is to appeal to the property exemplification theory, whereby instance A, [x, F, t], = instance B, [y, G, t'], iff x = y, F = G, and t = t'. However, as has been noted by Kim, if *this* is the condition that instances must meet, then the token identity view is a non-starter; as Kim states:

> Take any token psychological event under the property exemplification view. It will be an object having a psychological property, P, at time *t*. If this is also a token neurological event, there must be a neural property, N, such that the same object has N at *t*, and the events, the object's having P at *t* and its having N at *t*, are one and the same event. From the identity criterion of events as property exemplifications, it then follows that psychological property P = neural property N! On this view of events, then, token identities imply type identities, and token

> physicalism will collapse to type physicalism (2012: 175; for similar reasoning see Horgan 1981: 400-401).

A key thing to note here is the extent to which this argument relies on the fine-grained account of instances that Kim offers. We said above that instances are intuitively singular entities – they are particulars. Kim's account of the identity conditions of instances makes this a very fine-grained matter: instance A, [x, F, t], = instance B, [y, G, t'], iff x = y, F = G, and t = t'. Intuitively, it seems plausible that there are instances in this sense, very fine-grained entities whose identity conditions are this strict. Speaking metaphorically: it seems as if we can 'chop up' the world this finely.

So how should NRPists respond here? One way would be to try and deny that mental-physical property identity follows from mental-physical instance identity even if we chop up the world this finely. NRPists might do this by appealing to properties which stand in d/d, or a similar relation. Consider an instance of blackness; intuitively, such an instance is an exemplification of the property *being a colour* and the property *being black*. Moreover, it seems possible to individuate a fine-grained instance of blackness, along Kimian lines, that is also an instance of colour. If colour and blackness are not identical, then property identity doesn't follow from instance identity.

However, even if we grant the line of thought, it will not help the NRPist, because, as we have seen at some length above, mental properties are not determinables of physical properties; not in the original, strict sense, nor any looser, more general sense spelled out in terms of, say, specificity (see Sections 3.4 and 3.5). Thus, if mental instances are identical to physical instances, and instances are Kimian, then mental-physical property identity follows.

Thus, if by 'instance' we mean 'property exemplification' in Kim's sense, then NRP is false. There is only one way to respond here on behalf of NRP: deny that there are property instances in the Kimian sense. There are some NRPists who already do this – those who understand the token identity claim to be a claim only about mental and physical events understood in a more coarse-grained, Davidsonian (or quasi-Davidsonian) sense, where event, E, can be identical to E*, even though E and E* instantiate different properties. For example, E can instantiate the property of being a belief, and E* can instantiate the property of being a brain state. Events in this Davidsonian sense, the thought goes, are the proper subject of the token identity claim, not property instances in Kim's sense – the latter simply aren't part of the correct ontology: they don't exist.

I call this response the 'nuclear option'. It's nuclear because it denies the existence of an entire category of entity which intuitively exists; as we saw above – it does seem possible to 'chop up' the world as Kim suggests. Moreover, this nuclear response is being pursued purely in order to save a view in the philosophy of mind, NRP, many of whose proponents often frame the view by referring to instances (as we have seen above – see Section 3.6 and 3.7). But it is not clear why we should simply abandon the thought that we can chop up the world in Kim's fine-grained way just to save NRP.

There is a relevant historical parallel here. Recall our discussion of Davidson in Chapter 1 (Section 1.4). When Davidson first faced the charge of epiphenomenalism, the charge was framed in the following terms: if anomalous monism is true, then mental properties are causally irrelevant. In response, Davidson took the nuclear option: properties don't exist, he claimed (despite using the word 'property' frequently). This response, it was widely held, was not satisfactory, and interest in anomalous monism began to decline, in significant part as a result of Davidson going for the nuclear option of denying the existence of properties in order to respond to the charge of epiphenomenalism. Perhaps Davidson was right, perhaps properties don't exist; but talk of properties was too deeply embedded in the wider debate, so denying their existence ended up removing Davidson's view from that debate rather than persuading his opponents. In abandoning properties as ontological posits, Davidson, in effect, abandoned the debate he was engaged with.

We face a similar situation here with regard to those NRPists who will deny that there are property instances in the Kimian sense. Perhaps those NRPists are right: perhaps we should reject such ontological posits, but such a rejection comes with costs. First, we reject what seems to be eminently possible: that we can chop up the world in this fine-grained way. Second, those NRPists who simply want to abandon instances as an ontological category, i.e. to take the nuclear option, risk simply abandoning the debate, *à la* Davidson.

There is also something of a *tu quoque* worry here: if non-reductionists can help themselves to a denial of the existence of instances, then why can't reductionists help themselves to a denial of, say, realisation? That would certainly make life for RP a lot more straightforward. But, of course, that would not be a wise move on behalf of reductionism – it would be a nuclear option; it just seems like there is (or at least could be) realisation, in some sense or other, and it is such a significant part of the debate that charity alone demands an approach which is more, let's say: 'diplomatic'. Thus, RPists are well advised to try, instead, to accommodate realisation within a reductive framework; as indeed, we have tried to do above (see Section 1.5). And non-reductionists would be wise to try and accommodate fine-grained instances for the same reasons.

It should also be noted that NRPists can't simply make a brute identity claim about mental and physical instances, alongside an insistence that mental properties are not identical to physical properties. Because, aside from the obscurity of such a move, as we have seen, *ad nauseam*, brute trans-ordinal identity claims are not a viable option for physicalists.

Thus we can conclude that, given a Kimian understanding of 'instance', there is a general inference from an instance of F = an instance of G, to F = G (where F and G do not stand in the d/d relation in any sense of 'd/d' we have explored above).

This leaves the two proper parthood options. Let's consider each in turn.

Perhaps mental instances are proper parts of physical instances. However, a dilemma is faced here: either mental instances are physical, or mental instances are not physical.

First Horn: mental instances are physical, because they are identical to proper parts of physical instances, and those proper parts are physical. But if mental instances are physical, then the argument from instance to property identity can run, and NRP is false.

Second horn: mental instances are not physical, they are non-physical mental parts of physical instances. But this is inconsistent with physicalism: everything is physical, or depends on it. These non-physical mental instances aren't physical, *ex hypothesi*, so if they are to be consistent with physicalism then they must depend on the physical. But they are supposed to be *parts* of physical instances, and wholes, we are told, depend on their parts, so, according to this hypothesis, a whole range of physical instances depend on non-physical mental instances as parts. But if that's consistent with physicalism, then I'm a banana.[100]

[100] Indeed, this view looks like a variety of panpsychism – physical instances necessarily containing non-physical mental instances as parts.

Only the last option remains: that mental instances have only physical instances as proper parts. But if x has only physical proper parts, then x is physical, so if mental instances have only physical proper parts, then mental instances are physical instances, and the above argument from instance to property identity can run.

Thus, whichever one of the three options is taken, mental instances are physical instances, and mental properties must be physical properties, so reductionism follows.

We turn now to the second problem for NRP.

4.2 A Rylean ghost in Putnam's machine?

My argument in brief: mental instances are spatially extended, physicalism is inconsistent with extended non-physical entities–ghosts–so mental instances must be physical. If mental instances are physical, then mental properties are physical.

Allow me to sketch the general thought dogmatically to begin with: it should be agreed that both mental and physical instances have *spatial extension*; i.e. whatever else we want to say about such instances, we should agree that they occupy a volume of space. The kind of physical instances which are plausible candidates for realising mental instances all obviously have extension: c-fibre activations, Martian feet cavity inflations, and so on. If we accept that the physical instances that realise mental instances have location *and* extension, then we should accept that the mental instances are extended too. If it's obvious that Alex's c-fibre activation is extended, and we accept that c-fibre activation is realising Alex's pain, then we should accept that Alex's pain is extended too. If Alex's pain is extended, then it must be physical, otherwise it would be a kind Rylean ghost. Generalising the line of thought, all realised mental instances are physical, and the argument from instance to property identity can run.

How might one object to this line of reasoning? To answer that, let's lay out some of the key steps more formally:

The ghost argument

P1: All physical realiser instances are extended.

P2: If P1, then all realised mental instances are extended.

C1: All realised mental instances are extended.

P3: All extended instances are physical (there are no ghosts).

C2: All realised mental instances are physical

P4: If C2, then mental properties are physical properties.

C3: Mental properties are physical properties.

Now let's consider each premise in turn. I think any attempts to deny P1 are best met with an incredulous stare. But in the interests of charity, my reasoning in support of P1 is as follows.

One might deny that realiser instances are concrete at all, perhaps they are tropes: abstract particulars. But this is simply wrongheaded. Consider an instance of c-fibre activation, this is clearly not an abstract entity – perhaps it is a bundle of abstract particulars (assuming concreta can be so constructed), but it is not itself abstract.

A lot hangs here on what is meant by 'abstract' and 'concrete'. These are terms of art. Here's two very plausible conditions that I endorse:

x is *concrete* if x has a spatial location.

x is *abstract* only if x does not have a spatial location.

Given this understanding, we can see why it is obviously false that realising instances are abstract. One might mean something different by 'abstract' of course, but given any of the standard understandings of the term, realising instances are not abstract (see Rosen 2012).

Attempts to deny that realiser instances have spatial location are similarly wrongheaded. C-fibre activation instances are in brains, tiny cavity inflation instances are in Martian feet, and so on.

One could grant that realiser instances are concrete, and have spatial location, but deny that they have extension. There are examples of concrete instances which do not have extension, such as instances of being a geometric point in space, but these are very poor candidates for realising mental instances. The prime candidates invoked in discussions of specific cases of realisation involve such things as c-fibre activation instances or other biological or chemical instances, and it seems difficult to deny that *they* have extension. People sometimes talk in terms of microphysical instances, Shoemaker's MSE* instances, for example, but they have extension too.

Thus, there are very strong grounds for accepting P1.

Now consider P2: if all physical realiser instances are extended, then all realised mental instances are extended. There is a tradition in the philosophy of mind which views the idea that mental entities might have spatial location, never mind extension, with some suspicion. Here is a classic example from Malcolm:

> It is clear that a brain process has spatial location [but] consider an example of a sudden thought […] is this sudden thought literally inside my skull? I think that in our ordinary use of the terms 'thought' and 'thinking', we attach no meaning to the notion of determining the bodily location of a thought. We do not seriously debate whether someone's sudden thought occurred in his heart, his throat, or his brain (1970: 174).

However, Malcolm then suggests that 'if massive correlations were discovered between thoughts and brain processes then we might *begin* to locate thoughts in the head' (ibid). And this, indeed, is increasingly the position that we find ourselves in today. Moreover, if we are to be physicalists, as Malcolm was not, and endeavour to do justice to the claim that the mental is realised by the physical, then the need to locate mental instances becomes irresistible.

Perhaps the most obvious way to deny this is to insist that mental instances are abstract, their realisers concrete. This, indeed, would echo the original sense of 'realise'–what Putnam had in mind when first introducing 'realisation' into the lexicon in 'Minds and machines'.[101] Recall that Putnam stated that:

[101] Kim has a very instructive discussion of the history of 'realisation' (2010: 101-103).

> the ‘logical description’ of a Turing machine does not include any specification of the physical nature of [the machine’s internal] ‘states’ or indeed, of the *physical nature* of the whole machine. (Shall it consist of electronic relays, of cardboard, of human clerks sitting at desks, or what?) In other words, a given ‘Turing machine’ is an *abstract* machine which may be physically *realized* in an almost infinite number of different ways […] It is interesting to note that just as there are two possible descriptions of the behavior of a Turing machine–the engineer’s structural blueprint and the logician’s ‘machine table’–so there are two possible descriptions of human psychology. The ‘behavioristic’ approach […] aims at eventually providing a complete physicalistic description of human behaviour in terms which link up with chemistry and physics. This corresponds to the […] physicist’s description of a physically realized Turing machine. But it would also be possible to seek a more abstract description of human mental processes, in terms of ‘mental states’ (physical *realization*, if any, unspecified) (1960: 159-161; emphasis added).

And this general picture is at the heart of classic NRP: mental properties are defined by their causal roles – they are the property of having some property or other that plays the relevant causal role. Causal roles are abstract, but can be played by concreta; and in this way the mental can realise the physical.

But instances are by their nature concrete: an object instantiating a property at a time. And as all good physicalists know, one can be in time only by being in space (at least in the worlds physicalists care about). So even if mental properties are abstract (even if all properties are abstract), their instances must be concrete.

Thus, mental instances have spatial location. So, if one wants to deny that mental instances have extension, then the only option is to claim that mental instances are geometric points. But we can see how implausible this proposal is if we examine it in the light of the NRPist models we have considered.

According to classic NRP, there are only physical instances–those instances which ‘play the role’ of the second-order mental property. The second-order mental property’s instances just are those physical role-playing instances (this is the token identity view). Bolting on the claim that the second-order mental property has ‘geometric point’ type instances in addition to those physical role-playing instances, just to deny P2 of the ghost argument, would be a rather bizarre, *ad hoc* move.

According to the stand-alone account of d/d, there are mental instances and physical instances – the c-fibre activation instance is a specific way of being the pain instance. If there's a c-fibre activation instance, then there's a pain instance, in a way supposedly analogous to: if there's an instance of scarlet, then there's an instance of red. But consider the claim that an instance of scarlet, ${}_{I}S$, has extension but the instance of red, ${}_{I}R$, which ${}_{I}S$ specifies, is a geometric point. It is something of an understatement to say that it is far from clear how ${}_{I}S$ could specify ${}_{I}R$ if the latter is a geometric point and the former is extended. The scarlet instance will have a certain hue, saturation, and brilliance, but geometric points can't have a hue, saturation, and brilliance – they're geometric points! And all of this assumes that it makes sense to talk about pain or red instances being geometric points in the first place, of which we should be highly doubtful. So, if physical instances are specific ways of being mental instances, then P2 follows.

According to Pereboom's version of c/c, every mental instance is spatially coincident with a physical instance, so if physical instances are extended, then P2 follows.

According to Paul's version of c/c, every mental instance that has anything to do with causation is identical to a physical instance. Thus, for all the instances that really count, if the physical instance is extended, then the mental instance will be too. So P2 follows.

According to Yablo's version of s/s, mental events are realised by physical events in virtue in virtue of the essences of the former being a subset of the essences of the latter. Thus, on this view, the only entities relevant to P2 of the ghost argument are the events themselves which, by Yablo's account, stand in the d/d relation. So, for present purposes, Yablo's appeal to s/s is an appeal to d/d, and P2 follows in the manner we have already outlined above.[102]

According to Wilson and Shoemaker's version of s/s, mental instances are identical to physical instances, so if the latter are extended then so are the former. So P2 follows.

[102] We might ask after the instances of the mental *essences* themselves, as opposed to the events of which they are the essences, but according to Yablo mental essences aren't realised by the physical essences that necessitate them – instead, it is the events made up of those essences that stand in the realisation relation.

Thus, every variant of NRP that we have considered accepts that mental instances are extended.

This brings us to P3: all extended instances are physical (there are no ghosts). The main reason to endorse this premise is its empirical plausibility – there have not been any cases of people bumping into ghosts. But if there really were extended, non-physical instances in the world, then one would expect that someone somewhere would have empirically detected one by now.

Of course, one can always explain away a failure to detect such ghosts by this or that *ad hoc* move. Perhaps, it could be claimed, ghosts occupy all and only the same space as physical instances and aren't empirically detectable. But this is clearly not a plausible line for a physicalist to take. Imagine what the consequences would be if the possibility of undetectable non-physical extended entities were to be embraced in an effort to defend a physicalist view: ectoplasm, spirits from netherworld, djinns, chakras, and any manner of new age bunkum would be allowed into the material realm to run amok – for on what *physicalist* basis could we rule such things out? *This* version of physicalism allows for non-physical extended entities – so the sky's the limit!

Thus, any physicalist worth their materialist salt must exorcise ghosts from their theory; and all that remains is to appeal to the general inference from instance to property identity that we have already established, and which constitutes P4 of the ghost argument. This gives us our conclusion: mental properties are identical to physical properties.

4.3 Conclusion

We began by considering the state of play between RP and NRP, with one version of each appearing the most plausible. For NRP, classic non-reductionism offered both robust non-identity and an internal relation: causal role-playing. For RP, the kind of functionalist analysis offered by realiser functionalism, combined with the identification of mental properties with a disjunction of their realisers, looked to be the best bet, offering an internal relation, accommodating multiple realisation and autonomy, and the identification of mental and physical properties.

We then considered two general arguments to show that RP is to be preferred to NRP: the instance argument, and the ghost argument. For the former, we considered three options: mental instances are identical with physical instances, the former are proper parts of the latter, or the latter are the proper parts of the former. All three options, we saw, led to the view that mental instances are physical instances. We then established the validity of the inference from the identity of instances of F and G to the identity of F and G (except where F and G stand in d/d to each other). This allowed us to conclude that if mental instances are physical instances, then mental properties are physical properties.

We then turned to the ghost argument, which showed that mental instances must be extended, and all extended instances must be physical instances. This opened the way towards appealing to the general inference from instance to property identity. Thus, we again concluded that if mental instances are physical instances, then mental properties are physical properties.

This gives us our conclusion: if the choice is between reductive or non-reductive physicalism, then we should choose the former over the latter.

Bibliography

Antony, Louise M. (1991). The causal relevance of the mental. *Mind and Language* 6 (4):295-327.

Antony, Louise M. (1999). Multiple realizability, projectibility, and the reality of mental properties. *Philosophical Topics* 26 (1/2):1-24.

Antony, Louise M. (2003). Who's afraid of disjunctive properties? *Philosophical Issues* 13 (1):1-21.

Antony, Louise M. (2007). Everybody has got it: A defense of non-reductive materialism. In Brian P. McLaughlin & Jonathan D. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*. Blackwell, 143-149.

Armstrong, David M. (1968). *A Materialist Theory of the Mind*. Routledge.

Armstrong, David M. (1978). *Universals and Scientific Realism: A Theory of Universals*, Vol. Ii. Cambridge University Press.

Árnadóttir, Steinvör Thöll & Crane, Tim (2013). There is no exclusion problem. In Sophie C. Gibb & Rögnvaldur Ingthorsson (eds.), *Mental Causation and Ontology*. Oxford University Press, 248-265.

Batterman, Robert (2012). Intertheory relations in physics. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Csli, Stanford University.

Baker, Lynne Rudder (1993). Metaphysics and mental causation. In John Heil & Alfred R. Mele (eds.), *Mental Causation*. Oxford University Press, 75-96.

Baker, Lynne Rudder (1997). Why constitution is not identity. *Journal of Philosophy* 94 (12):599-621.

Baker, Lynne Rudder (2000). *Persons and Bodies: A Constitution View*. Cambridge University Press.

Baker, Lynne Rudder (2002). The ontological status of persons. *Philosophy and Phenomenological Research* 65 (2):370-388.

Baker, Lynne Rudder (2007). *The Metaphysics of Everyday Life: An Essay in Practical Realism*. Cambridge University Press.

Baker, Lynne Rudder (2009). Non-reductive materialism. In Brian P. McLaughlin, Ansgar Beckermann & Sven Walter (eds.), *The Oxford Handbook of Philosophy of Mind*. Oxford University Press, 109-127.

Bedau, Mark, & Humphreys, Paul (2008). *Emergence: Contemporary Readings in Philosophy and Science*. MIT Press.

Bennett, Karen (2003). Why the exclusion problem seems intractable and how, just maybe, to tract it. *Noûs* 37 (3):471-97.

Bennett, Karen (2008). Exclusion again. In Jakob Hohwy & Jesper Kallestrup (eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation*. Oxford University Press, 280-304.

Bickle, John (2008). Multiple realizability. *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Csli, Stanford University.

Blackburn, Simon W. (1984). Supervenience revisited. In Ian Hacking (ed.), *Exercises in Analysis: Essays by Students of Casimir Lewy*. Cambridge University Press, 59-74.

Blackburn, Simon (2008). *The Oxford Dictionary of Philosophy*. Oxford University Press.

Block, Ned (1990). Can the mind change the world? In Hilary Putnam & George Boolos (eds.), *Meaning and Method: Essays in Honour of Hilary Putnam*. Cambridge University Press, 137-170.

Block, Ned (1996). Anti-reductionism slaps back: Mental causation, reduction and supervenience. *Philosophical Perspectives* 11:107-132.

Block, Ned (1997). Anti-reductionism slaps back. *Noûs* 31(11):107-132.

Block, Ned & Stalnaker, Robert (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review* 108 (1):1-46.

Bontly, Thomas D. (2005). Proportionality, causation, and exclusion. *Philosophia* 32 (1-4):331-348.

Boring, E. G. (1933). *The Physical Dimensions of Consciousness*. The Century Co.

Borst, Clive Vernon (1970). *The Mind-Brain Identity Theory: A Collection of Papers*. The Macmillan Press Ltd.

Bradley, F. H. (1893). *Appearance and Reality*. Clarendon Press.

Broad, C. D. (1925). *The Mind and its Place in Nature*. Routledge and Kegan Paul.

Brooks, D. H. M. (1994). How to perform a reduction. *Philosophy and Phenomenological Research* 54 (4):803-14.

Bunzl, Martin (1979). Causal overdetermination. *Journal of Philosophy* 76 (3):134-150.

Carey, Brandon (2011). Overdetermination and the exclusion problem. *Australasian Journal of Philosophy* 89 (2):251 - 262.

Carnap, R., (1932). Psychologie in Physikalischer Sprache. *Erkenntnis*, 3: 107–142. English translation in A.J. Ayer (ed.) (1959). *Logical Positivism*. Free Press.

Chalmers, David J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, David J. (1999). Materialism and the metaphysics of modality. *Philosophy and Phenomenological Research* 59 (2):473-96.

Christensen, J. & Kallestrup, J. (2012). Counterfactuals and downward causation: a reply to Zhong. *Analysis* 72 (3):513-517.

Clapp, Lenny (2001). Disjunctive properties. *Journal of Philosophy* 98 (3):111 - 136.

Crane, Tim & Mellor, D. H. (1990). There is no question of physicalism. *Mind* 99 (394):185-206.

Crisp, Thomas M., & Warfield, Ted A. (2001). Kim's master argument. *Noûs* 35 (2):304–316.

Davidson, Donald (1970). Mental events. In L. Foster & J. W. Swanson (eds.), *Experience and Theory*. Humanities Press, 79-101.

Davidson, Donald (1980). *Essays on Actions and Events*. Oxford University Press.

Davidson, Donald (1993). Thinking causes. In John Heil & Alfred R. Mele (eds.), *Mental Causation*. Oxford University Press, 3-18.

Dowe, Phil (2000). *Physical Causation*. Cambridge University Press.

Ehring, Douglas E. (1996). Mental causation, determinables, and property instances. *Noûs* 30 (4):461-80.

Fair, David (1979). Causation and the flow of energy. *Erkenntnis* 14 (3):219 - 250.

Feigl, Herbert (1958). The 'mental' and the 'physical'. *Minnesota Studies in the Philosophy of Science* 2:370-497.

Fine, Kit (2003). The non-identity of a material thing and its matter. *Mind* 112 (446):195-234.

Fodor, Jerry A. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese* 28 (2):97-115.

Fodor, Jerry A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.

Fodor, Jerry A. (1997). Special sciences: Still autonomous after all these years. *Philosophical Perspectives* 11 (s11):149-63.

Funkhouser, Eric (2006). The determinable-determinate relation. *Noûs* 40 (3):548–569.

Garrett, Brian J. (1998). Pluralism, causation, and overdetermination. *Synthese* 116 (3):355-78.

Gibb, Sophie C. (2006). Why Davidson is not a property epiphenomenalist. *International Journal of Philosophical Studies* 14 (3):407 – 422.

Gibb, Sophie C. (2013). Introduction. In Sophie C. Gibb & Rögnvaldur Ingthorsson (eds.), *Mental Causation and Ontology*. Oxford University Press. 1-18.

Gillett, Carl (2002). The dimensions of realization: A critique of the standard view. *Analysis* 62 (4):316-323.

Gillett, Carl (2003). The metaphysics of realization, multiple realizability, and the special sciences. *Journal of Philosophy* 100 (11):591-603.

Gillett, Carl (2007). A mechanist manifesto for the philosophy of mind: A third way for functionalists. *Journal of Philosophical Research* 32:21-42.

Goldman, Alvin I. (1969). The compatibility of mechanism and purpose. *Philosophical Review* 78 (October):468-82.

Gozzano, Simone & Hill, Christopher S. (eds.) (2012). *New Perspectives on Type Identity: The Mental and the Physical*. Cambridge University Press.

Hall, Ned (2004). Two concepts of causation. In John Collins, Ned Hall & Laurie Paul (eds.), *Causation and Counterfactuals*. The MIT Press, 225-276.

Haug, Matthew C. (2010). Realization, determination, and mechanisms. *Philosophical Studies* 150 (3):313-330.

Hempel, Carl G. (1980). Comments on Goodman's ways of worldmaking. *Synthese* 45 (2):193 - 199.

Hempel, Carl G. & Oppenheim, Paul (1948). Studies in the logic of explanation. *Philosophy of Science* 15 (2):135-175.

Heil, John (1992). *The Nature of True Minds*. Cambridge University Press.

Hill, Christopher S. (1991). *Sensations: A Defense of Type Materialism*. Cambridge University Press.

Horgan, Terence E. (1981). Token physicalism, supervenience, and the generality of physics. *Synthese* 49 (December):395-413.

Horgan, Terence E. (1993). From supervenience to superdupervenience: Meeting the demands of a material world. *Mind* 102 (408):555-86.

Horgan, Terence E. (1997). Kim on mental causation and causal exclusion. *Philosophical Perspectives* 11 (s11):165-84.

Howell, Robert J. (2009). Emergentism and supervenience physicalism. *Australasian Journal of Philosophy* 87 (1):83 – 98.

Huxley, T. H. (1866). *Lessons in Elementary Physiology*. Macmillan and Co.

Huxley, T. H. (1874). *Science and Culture and Other Essays*. Macmillan and Co.

Jackson, Frank. (2012). In defence of the identity theory, mark I. In Hill Christopher & Gozzano Simone (eds.), *New Perspectives on Type Identity: The Mental and the Physical*. Cambridge University Press, 150-166.

Jackson, Frank & Pettit, Philip (1990). Program explanation: A general perspective. *Analysis* 50 (2):107-17.

Kallestrup, Jesper (2006). The causal exclusion argument. *Philosophical Studies* 131 (2):459-85.

Kim, Jaegwon (1966). On the psycho-physical identity theory. *American Philosophical Quarterly* 3 (July):227-35.

Kim, Jaegwon (1973a). Causation, nomic subsumption, and the concept of event. *Journal of Philosophy* 70 (8):217-236.

Kim, Jaegwon (1973b). Causes and counterfactuals. *Journal of Philosophy* 70 (17):570-572.

Kim, Jaegwon (1974). Noncausal connections. *Noûs* 8 (1):41-52.

Kim, Jaegwon (1977). Causation, emphasis, and events. *Midwest Studies in Philosophy* 2 (1):100-103.

Kim, Jaegwon (1979). Causality, identity and supervenience in the mind-body problem. *Midwest Studies in Philosophy* 4 (1):31-49.

Kim, Jaegwon (1984a). Epiphenomenal and supervenient causation. *Midwest Studies in Philosophy* 9 (1):257-70.

Kim, Jaegwon (1984b). Supervenience and supervenient causation. *Southern Journal of Philosophy Supplement* 22 (S1):45-56.

Kim, Jaegwon (1988). Supervenience for multiple domains. *Philosophical Topics* 16 (1):129-50.

Kim, Jaegwon (1989a). The myth of non-reductive materialism. *Proceedings and Addresses of the American Philosophical Association* 63 (3):31-47.

Kim, Jaegwon (1989b). Mechanism, purpose, and explanatory exclusion. *Philosophical Perspectives* 3:77-108.

Kim, Jaegwon (1990a). Supervenience as a philosophical concept. *Metaphilosophy* 21 (1-2):1-27.

Kim, Jaegwon (1990b). Explanatory exclusion and the problem of mental causation. In Enrique Villanueva (ed.), *Information, Semantics, and Epistemology*. Blackwell, 36-56.

Kim, Jaegwon (1992a). The nonreductivist's trouble with mental causation. In John Heil & Alfred R. Mele (eds.), *Mental Causation*. Oxford University Press, 189-210.

Kim, Jaegwon (1992b). ‘Downward causation’ in emergentism and nonreductive physicalism. In Ansgar Beckermann, Hans Flohr & Jaegwon Kim (eds.), *Emergence or Reduction? Prospects for Nonreductive Physicalism*. De Gruyter, 117-138.

Kim, Jaegwon, (1992c). Multiple realisation and the metaphysics of reduction. *Philosophy and Phenomenological Research* 52 (1):1-26.

Kim, Jaegwon (1993a). *Supervenience and Mind.* Cambridge University Press.

Kim, Jaegwon (1993b). Mental causation in a physical world. In Enrique Villanueva (ed.), *Science and Knowledge*. Ridgeview, 27-50.

Kim, Jaegwon (1995). Mental causation: What? Me worry? *Philosophical Issues* 6:123-151.

Kim, Jaegwon (1996). *Philosophy of Mind* (First Edition). Westview Press.

Kim, Jaegwon (1997). Does the problem of mental causation generalize? *Proceedings of the Aristotelian Society* 97 (3):281-97.

Kim, Jaegwon (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. MIT Press.

Kim, Jaegwon (1999a). Supervenient properties and micro-based concepts: A reply to Noordhof. *Proceedings of the Aristotelian Society* 99 (1):115-118.

Kim, Jaegwon (1999b). Hempel, explanation, metaphysics. *Philosophical Studies* 94 (1-2):1-20.

Kim, Jaegwon (2000). How can my mind move my Limbs? Mental causation from Descartes to contemporary physicalism. *Philosophic Exchange* 30:5-16.

Kim, Jaegwon (2001). Mental causation and consciousness: The two mind-body problems for the physicalist. In Carl Gillett & Barry M. Loewer (eds.), *Physicalism and its Discontents*. Cambridge University Press, 271-283.

Kim, Jaegwon (2002). Responses to comments on *Mind in a Physical World*. *Philosophy and Phenomenological Research* 65 (3):671-680.

Kim, Jaegwon (2003). Blocking causal drainage and other maintenance chores with mental causation. *Philosophy and Phenomenological Research* 67 (1):151-176.

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

Kim, Jaegwon (2006). *Philosophy of Mind* (Second Edition). Westview Press.

Kim, Jaegwon (2007a). Causation and mental causation. In Brian P. McLaughlin & Jonathan D. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*. Blackwell, 227-242.

Kim, Jaegwon (2007b). The causal efficacy of consciousness. In Max Velmans & Susan Schneider (eds.), *The Blackwell Companion to Consciousness*. Blackwell, 406-417.

Kim, Jaegwon (2009). Mental causation. In Brian P. McLaughlin, Ansgar Beckermann & Sven Walter (eds.), *The Oxford Handbook of Philosophy of Mind*. Oxford University Press, 29-52.

Kim, Jaegwon (2010). Thoughts on Sydney Shoemaker's physical realization. *Philosophical Studies* 148 (1):101 - 112.

Kim, Jaegwon (2011). *Philosophy of Mind* (Third Edition). Westview Press.

Kirk, Robert (2013). *The Conceptual Link From Physical to Mental*. Oxford University Press.

Klein, Colin (2009). Reduction without Reductionism: A Defence of Nagel on Connectability. *Philosophical Quarterly* 59 (234):39 - 53.

Kripke, Saul A. (1971). Identity and necessity. In Milton K. Munitz (ed.), *Identity and Individuation*. New York University Press, 135-164.

Kroedel, Thomas (2008). Mental causation as multiple causation. *Philosophical Studies* 139 (1):125-143.

Kroedel, Thomas (2013). Dualist Mental Causation and the Exclusion Problem. *Noûs* 47 (3).

Law, Jonathan, & Smullen, John (2008). *A Dictionary of Finance and Banking* (Fourth Edition, revised). Oxford University Press.

Levine, Joseph (1993). On leaving out what it's like. In Martin Davies & Glyn W. Humphreys (eds.), *Consciousness: Psychological and Philosophical Essays*. Blackwell, 543-557.

Levine, Joseph (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford University Press.

Lewes, G. H. (1875). *Problems of Life and Mind*, Vol. 2. Kegan Paul, Trench, Turbner, & Co.

Lewis, David (1966). An argument for the identity theory. *Journal of Philosophy* 63 (2):17-25.

List, Christian, & Menzies, Peter (2009). Nonreductive physicalism and the limits of the exclusion principle. *Journal of Philosophy* 106 (9):475-502.

Lowe, E. J. (1983). Instantiation, identity and constitution. *Philosophical Studies* 44 (1):45 - 59.

Lowe, E. J. (1995). Coinciding Objects: In Defence of the 'Standard Account'. *Analysis* 55 (3):171 - 178.

Lowe, E. J. (1998). *The Possibility of Metaphysics: Substance, Identity, and Time*. Oxford University Press.

Lowe, E. J. (2000). Causal closure principles and emergentism. *Philosophy* 75 (294):571-586.

Loewer, Barry (2002). Comments on Jaegwon Kim's mind and the physical world. *Philosophy and Phenomenological Research* 65 (3):655–662.

Loewer, Barry (2007). Mental causation, or something near enough. In Brian P. McLaughlin & Jonathan D. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*. Blackwell, 243-265.

Lycan, William G. (1986). Moral facts and moral knowledge. *Southern Journal of Philosophy* 24 (S1):79-94.

Lynch, Michael P. & Glasgow, Joshua (2003). The impossibility of superdupervenience. *Philosophical Studies* 113 (3):201-221.

MacDonald, Cynthia, & MacDonald, Graham F. (1986). Mental causes and explanation of action. *Philosophical Quarterly* 36(143):145-158.

Macdonald, Cynthia, & Macdonald, Graham (eds.), *Philosophy of Psychology: Debates on Psychological Explanation*. Blackwell.

Malcolm, Norman (1968). The conceivability of mechanism. *Philosophical Review* 77 (January):45-72.

Malcolm, Norman (1970). Scientific materialism and the identity theory. In C. V. Borst (ed.) (1970) *The Mind-Brain Identity Theory: A Collection of Papers*. The Macmillan Press Ltd, 171-180.

Marras, Ausonio (2002). Kim on reduction. *Erkenntnis* 57 (2):231-57.

Marras, Ausonio (2003). Methodological and ontological aspects of the mental causation problem. In Sven Walter & Heinz-Dieter Heckmann (eds.), Physicalism and Mental Causation. Imprint Academic, 243-264.

Martin, Michael W. (1971). On the conceivability of mechanism. *Philosophy of Science* 38 (March):79-86.

Maslen, Cei, Horgan, Terry, & Daly, Helen (2009). Mental Causation. In Helen Beebee, Christopher Hitchcock & Peter Menzies (eds.), *The Oxford Handbook of Causation*. Oxfrord University Press, 523-553.

Maund, Barry (2012). Color. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Csli, Stanford University.

McGrath, M. (1998). Proportionality and mental causation: A fit? *Philosophical Perspectives* 12 (S12):167-176.

McLaughlin, Brian P. (1992). The rise and fall of british emergentism. In Ansgar Beckermann, Hans Flohr & Jaegwon Kim (eds.), *Emergence or Reduction? Prospects for Nonreductive Physicalism*. De Gruyter, 49-93.

McLaughlin, Brian P. (2001). In defense of new wave materialism: A response to Horgan and Tienson. In Carl Gillett & Barry M. Loewer (eds.), *Physicalism and its Discontents*. Cambridge University Press, 319-330.

McLaughlin, Brian P. (2007). On the limits of *a priori* physicalism. In Brian P. McLaughlin & Jonathan D. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*. Blackwell, 200-223.

McLaughlin, Brian P., Beckermann, Ansgar & Walter, Sven (eds.) (2009). *The Oxford Handbook of Philosophy of Mind*. Oxford University Press.

McLaughlin, Brian & Bennett, Karen (2014). Supervenience. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Csli, Stanford University.

Melnyk, Andrew (2003). *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge: Cambridge University Press.

Melnyk, Andrew (2008). Can physicalism be non-reductive? *Philosophy Compass* 3 (6):1281-1296.

Merrick, Trenton (2001). *Objects and Persons*. Clarendon Press.

Mill, J. S. (1843). *System of Logic*. Longmans, Green, Reader, and Dyer.

Moore, G. E. (1919). External and Internal Relations. *Proceedings of the Aristotelian Society* 20:40- 62.

Nagel, Ernest (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*. Harcourt, Brace & World.

Ney, Alyssa (2007). Can an appeal to constitution solve the exclusion problem? *Pacific Philosophical Quarterly* 88 (4):486–506.

O’Connor, T. & Wong, H. Y. (2012). Emergence. *Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Csli, Stanford University.

Oppenheim, Paul & Putnam, Hilary (1958). Unity of Science as a Working Hypothesis. In Herbert Feigl, Michael Scriven & Grover Maxwell (eds.). University of Minnesota Press, 3-35.

Paul, L. A. (2007). Constitutive Overdetermination. In J. K. Campbell, M. O'Rourke & H. S. Silverstein (eds.), *Causation and Explanation*. Mit Press, 265-290.

Papineau, David (2013). Causation is macroscopic but not irreducible. In Sophie C. Gibb & Rögnvaldur Ingthorsson (eds.), *Mental Causation and Ontology*. Oxford University Press. 126-152.

Pereboom, Derk (2002). Robust nonreductive materialism. *Journal of Philosophy* 99 (10):499-531.

Pereboom, Derk (2011). *Consciousness and the Prospects of Physicalism*. Oxford University Press.

Place, U. T. (1956). Is consciousness a brain process? *British Journal of Psychology*, 47: 44–50.

Prior, Arthur N. (1949). Determinables, determinates and determinants. *Mind* 58 (229):1-20.

Putnam, Hilary (1960). Minds and machines. In Sidney Hook (ed.), *Dimensions of Mind*. New York University Press, 57-80.

Putnam, Hilary (1967). Psychological predicates. In W.H. Capitan & D.D. Merrill (eds.), *Art, Mind, and Religion*. Pittsburgh University Press, 37-48.

Putnam, Hilary (1970). On properties. In N. Rescher *et al*. (eds.), *Essays in honor of Carl G. Hempel*. Reide.

Richardson, Robert C. (1979). Functionalism and reductionism. *Philosophy of Science* 46 (4):533-58.

Robb, David, & Heil, John (2013). Mental Causation. *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Csli, Stanford University.

Ryle, Gilbert (1949). *The Concept of Mind*. Hutchinson and Co.

Salmon, Wesley C. (1998). *Causality and Explanation*. Oxford University Press.

Sanford, David H. (2008). Determinates vs. determinables. *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Csli, Stanford University.

Hardin, C. L. (1988). *Color for Philosophers*. Hackett.

Schiffer, Stephen R. (1987). *Remnants of Meaning*. MIT Press.

Körner, S. & Searle, J. (1959). Symposium: On determinables and resemblance. *Aristotelian Society Supplementary Volume* 33:125 - 158.

Shapiro, Lawrence A. & Polger, Thomas W. (2012). Identity, variability, and multiple realization in the special sciences. In Hill Christopher & Gozzano Simone (eds.), *New Perspectives on Type Identity: The Mental and the Physical*. Cambridge University Press, 264-287.

Shoemaker, Sydney (2001). Realization and mental causation. In Carl Gillett & Barry M. Loewer (eds.), Physicalism and its Discontents. Cambridge University Press, 23-33.

Shoemaker, Sydney (2007). *Physical Realization*. Oxford University Press.

Shoemaker, Sydney (2013). Physical realization without preemption. In Sophie C. Gibb & Rögnvaldur Ingthorsson (eds.), *Mental Causation and Ontology*. Oxford University Press, 35-56.

Sider, Theodore (2003). What's so bad about overdetermination? *Philosophy and Phenomenological Research* 67 (3):719 - 726.

Sklar, Lawrence (1967). Types of inter-theoretic reduction. *British Journal for the Philosophy of Science* 18 (2):109-124.

Sklar, Lawrence (2008). Philosophy of statistical mechanics. *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Csli, Stanford University.

Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review* 68 (April):141-56.

Smart, J. J. C. (2007). The identity theory of mind. *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Csli, Stanford University.

Tye, M. (1995). What what it's like is really like. *Analysis* 55 (2):125-126.

Unger, Peter (1977). The uniqueness in causation. *American Philosophical Quarterly* 14 (3):177 - 188.

van Riel, Raphael (2014). *The Concept of Reduction*. Springer.

Walter, Sven (2007). Determinables, determinates, and causal relevance. *Canadian Journal of Philosophy* 37 (2):217-244.

Weisberg, Michael, Needham, Paul, & Hendry, Robin (2011). Philosophy of chemistry. *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Csli, Stanford University.

Whittle, Ann (2007). The co-instantiation thesis. *Australasian Journal of Philosophy* 85 (1):61 – 79.

Wilson, Jessica M. (1999). How superduper does a physicalist supervenience need to be? *Philosophical Quarterly* 50 (194):33-52.

Wilson, Jessica M. (2009). Determination, realization and mental causation. *Philosophical Studies* 145 (1):149 - 169.

Wilson, Jessica M. (2011). Non-reductive realization and the powers-based subset strategy. *The Monist* (Issue on Powers) 94 (1):121-154.

Worley, Sara (1997). Determination and mental causation. *Erkenntnis* 46 (3):281-304.

Woodward, James (2011) Scientific explanation. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Csli, Stanford University.

Yablo, Stephen (1987). Identity, essence, and indiscernibility. *Journal of Philosophy* 84 (6):293-314.

Yablo, Stephen (1992). Mental causation. *Philosophical Review* 101 (2):245-280.

Yablo, Stephen (1997). Wide causation. *Philosophical Perspectives* 11 (11):251-281.

Yablo, Stephen (2000). Seven habits of highly effective thinkers. In Bernard Elevitch (ed.), *Proceedings of the Twentieth World Congress of Philosophy*, Vol. 9. Philosophy Documentation Center. 35-45.

Yablo, Stephen (2001). Superproportionality and mind-body relations. *Theoria* 16 (40):65-75.

Zhong, Lei (2011). Can Counterfactuals Solve the Exclusion Problem? *Philosophy and Phenomenological Research* 83 (1):129-147.

Zimmerman, Dean (2002). The constitution of persons by bodies: A critique of Lynne Rudder Baker's theory of material constitution. *Philosophical Topics* 30 (1):295-338.