

Name: Christopher Fisher

Student Number: 0605784J

Supervisor: Prof. Mark Leon

Title: 'The Place of Mental Properties in a Physical World'

Table of Contents

Introduction

1. Chapter One: Setting the Scene

1.1. The Positions and Their Arguments

1.2. A Hidden Premise

2. The Traditional Problem with Non-Reductive Materialism: Mental Causation

3. Chapter Three: Redefining the Landscape

3.1. Strong Property Dualism, Zombies and Substance Dualism.

3.1.1. Zombies, Modality and the Bearers of Mental Properties.

3.1.2. Zombies and Ghosts.

3.2. Supervenience and Non-Reductive Physicalism.

3.2.1. Alternate Necessities for Supervenience.

3.2.2. Explaining Supervenience.

3.3. Token Identity and Non-Reductive Physicalism.

3.4. A New Spectrum.

4. Chapter Three: An Alternative Physicalism.

4.1. Heil's View.

4.1.1. The Replaceability of (P).

4.1.2. Substance, Property and Truthmaking.

4.2. Immediate Benefits of Heil's View.

4.2.1. Mental Causation and Zombies.

4.3. The Avoidance of Eliminativism.

4.4. Mental Attribute Agreement.

4.5. Heil on Consciousness and Qualia.

4.5.1. Categorical and Dispositional Properties.

4.5.2. Representationalism.

4.5.3. The Benefits of Representationalism and Heil's Attenuated Version of it.

4.6. The Problem of Subjectivity and Further Trouble for Non-Reductive Physicalism

Conclusion

Bibliography

Introduction

In describing the mind's place in the physical world, philosophers have produced a diversity of views. On the far right of the spectrum we find substance dualists, who think that there is in addition to physical substance, mental substance. A little to the left strong property dualists claim that while there is only physical substance, there are in addition to physical properties, mental properties. These two views broadly represent accounts of mind on which physicalism is explicitly considered as false. Among those who think physicalism is true, are reductionists and non-reductionists. The former sit on the far left, claiming that there is only physical substance and that mental properties *just are* physical properties – the latter claim based on the purported success of one or another reductive procedure. Non-reductionists are at the center; rejecting the possibility of any reductive procedure, claiming that mental properties are distinct from physical ones, while resisting anti-physicalist claims about mental properties and substance. Throughout this paper, I will refer to the views on this spectrum as 'the traditional views.'

As they appear in the formulation of these positions, 'substance' and 'property' are *ontological categories*. This is just to say that each does more than indicate a different way of describing reality, each also picks out features *of reality*. (Thomasson)¹ So, when these positions endorse or deny a distinction formulated in terms of these categories, I will take that distinction to be making a claim about reality. In this way substance dualists, strong property dualists and non-reductionists all seem to claim that there is a distinction between the mental and the physical that is grounded in some sort of distinction in the world. Reductionists of course deny any sort of distinction.

In the first chapter of this paper I will set out each of these positions in more detail and argue that philosophers who find themselves somewhere on this spectrum are motivated by a tacit acceptance of the following premise:

(P): When a predicate or concept applies truly to an object, it does so in virtue of designating a [genuine] property possessed by that object and by every object to which the predicate truly applies (or would apply). (Heil and Robb 2003 p 176)

¹ Ontological Categories and How to Use Them, Amie L. Thomasson, <http://ejap.louisiana.edu/EJAP/1997.spring/thomasson976.html>

I will go on to argue that (P) carries with it a certain view about the relationship between reality and our descriptions of it as part of a larger metaphysical ‘big picture’. (Heil 2012 p. 11) I will also argue that (P) limits our choices regarding accounts of mind to the traditional views. This is problematic on the face of it, because at least three of the four traditional positions seem inherently problematic – namely, reductionist physicalism and non-physicalism of either form. I will take it for granted in this paper that reduction is likely to fail but that we still retain a bias towards a physicalist account of mental phenomena given the ubiquity of physicalist accounts of other higher-level phenomena. This leaves us with only non-reductionist brands of physicalism. These are problematic as well, however. This form of physicalism has notoriously been challenged on its ability to make sense of mental causation as well on difficulties in accounting for the apparently tight relationship between the mental and the physical, while preserving its *non-reductionist* credentials. Many have decided to persevere with non-reductionist physicalism, nonetheless, thanks to its open affirmation of physicalism and denial of reduction. It seems like it gives us everything we want.

In Chapter two, I will argue that the problem of mental causation as it arises for non-reductionists provides strong reason to reject their view. The argument of chapter two will consist of a critical discussion and broad defense of Jaegwon Kim’s critical literature regarding non-reductive accounts of mind and mental causation. I will argue in line with Kim that the non-reductionist will either need to endorse a view which claims that mental properties have causal powers that are ‘over and above’ the physical, or that the causal powers of mental properties are identical to the causal powers of the physical properties on which they depend. If the non-reductionist takes the first of these views he will compromise on his apparent physicalist credentials. If he takes the second view, there no longer seems to be any interesting sense in which the mental is distinct from the physical.

In Chapter three, I will argue that the philosophical landscape as presented above is mistaken. A core claim of this chapter will be that non-reductionist accounts of mind are unavailable because they are incoherent. I will argue that however we characterize non-reductive accounts, they seem either to lose their non-reductive credentials or their physicalist credentials. En route to this conclusion, I will show as well that there is no substantive distinction between strong property dualist accounts and substance dualist accounts. My strategy in chapter three will be to ask first how non-reductive physicalists could distinguish themselves from strong property dualists. I will consider two possibilities: The first is to claim that strong property dualists have a certain relationship to dualism about substances that non-reductionists do not: either strong property dualists cannot avoid dualism about substance or they are outright committed to it. The second way is to claim that the non-reductionist has

at his disposal a supervenience relation that characterizes the dependence between mental and physical differently to how the strong property dualist does. The difference here is often characterized such that non-reductionists appeal to a supervenience relation that tethers the mental to the physical more tightly than the strong property dualist's supervenience relation does.

With regard to the first way, I will argue that there is a strong case for both the claim that the strong property dualist can't avoid substance dualism and the claim that he is committed to it. I will present two separate arguments for each of these claims. Both arguments will center upon the strong property dualist's commitment to 'philosophical zombies'. I will then argue that the best way for the non-reductionist to avoid this line of argument is to appeal to a form of supervenience that characterizes the relation between mental and physical such that the possibility of zombies is ruled out. This is usually done through characterizing the 'modal strength' of the non-reductionist's supervenience relation such that it contrasts with the strong property dualist's appeal to supervenience. I will argue, that there is no way to characterize the modal strength of supervenience that coheres with the core claims of the non-reductionist. The result here will be the incoherence of non-reductive accounts in so far as they appeal to a supervenience relation.

The outcome of the foregoing discussion should already have begun to cast doubt on any strategy that appeals to the supervenience relation to further any non-reductionist cause. So, the second way of distinguishing strong property dualist and non-reductive accounts should already look bleak. But I will go on to consider a more traditional way of undermining the non-reductionists appeal to supervenience. Whatever the supervenience relation appealed to by the non-reductionist, it seems to require an explanation that makes *clear* that it is suitable for securing physicalism. I will argue that there is no substantive explanation of the supervenience such that it establishes its ability to secure physicalism that does not entail reduction. Moreover, I will show that the supervenience relation cannot be considered as brute and unexplainable and at the same time suitable to secure physicalism.

If my arguments in chapter two and three are sound, it would mean that at least we ought to reject non-reductive accounts of mind. This makes our limitation to the traditional views all the more troubling because we seem to be left with some rather extreme views as alternatives. Philosophers who reject my bias towards physicalism and against reductive forms of it, however, might take the considerations of this chapter to strengthen a *prima facie* case for either non-physicalism or reduction. Indeed, nothing I say in this paper will rule out reduction or non-physicalism. However, whether one accepts this bias or not, an important factor in

deciding where to place one's allegiance among the traditional positions is the apparent lack of an alternative physicalist account of mind.

In my fourth chapter I will suggest such an alternative. This will involve, most importantly, showing that we can get on without (P) and the metaphysical picture that comes along with it. This will lift the restriction to the traditional views. I will use as the basis of this chapter John Heil's account of the mind. He suggests an account of mind on which mental predicates refer to particular modes of complex physical substances. Heil is an 'irrealist' about mental properties in so far as he takes himself to be saying that there *are only* physical properties, strictly speaking. He thinks that whatever distinction there is between mental and physical, it consists purely in a distinction in concept, not in the world. In this way, Heil's view is neither like non-physicalists nor like non-reductionists. But it is not intended to be a form of reduction either. (Heil 2012 p. 1) He writes:

This is not a claim you could translate or analyze talk of trees into talk of electrons, quarks, fields. It is not the claim that you can replace biological taxonomies, concepts or terms, with taxonomies, concepts or terms at home in fundamental physics." (Heil 2012 p6)

Importantly, Heil does not take his irrealism about mental properties to entail eliminativism about mental states or mental vocabulary. He thinks that mental predicates still successfully refer to features of the mind independent world; *and* he thinks that vocabulary that deals in mental concepts is indispensable.

By avoiding allegiance to the traditional views, I will argue that Heil's view avoids the traditional problems as well. In particular, if Heil's view is correct, he is able to stop the problem of mental causation before it even arises and he is able to avoid the possibility of philosophical zombies. Heil's ability to avoid these problems, while not decisive in favor of his view, I think, makes it worthy of further exploration and worth the effort of trying to resolve *its* problems. Indeed, Heil's view has (quite literally) a lot of explaining to do. Having a commitment to (P) and distinct mental properties, while perhaps problematic in some ways, generally brings with it a great deal of the theoretical benefit. Due to their commitment to (P) and distinct mental properties, traditional views seem to have a great deal of explanatory power. They seem to have a ready made explanation as to how it is that at least some of our mental discourse is true – mental predicates refer to distinct mental properties. They also have an advantage in explaining how distinct items, sometimes with very different physical attributes, can agree in *mental* attribute – there is a common mental feature to each item that common mental predicates pick out. Traditional non-physicalist views also have on their side

highly intuitive explanatory gaps between physical properties and mental ones – particularly, so called ‘phenomenal properties’ of conscious experience. Philosophers have long been puzzled as to how it *could* be that the phenomenal character of experience is consistent with a world that has only physical properties. To the extent that these gaps cannot be bridged, Heil’s account will face difficulties.

Heil doesn’t bite any bullets here. He thinks that his view is not eliminativist because on it mental predicates still successfully refer to aspects of the mind independent world. He attempts answer to problem of attribute agreement by claiming that generality is a feature of our representations of reality, not necessarily a feature of reality itself. Generally applicable predicates, he says, do not require truthmakers that are *themselves* general. I will argue that whatever the virtues of Heil’s responses here, their ultimate success rests on Heil’s ability to account for the phenomenal properties of experience in his austere ontology. Heil makes two independent attempts to do this. The first is by appeal to a position on the metaphysical nature of properties. Heil claims that there is in fact no distinction between the categorical and dispositional natures of properties. For Heil, all properties are at once categorical and dispositional. He takes this to be a potential solution for the problem of qualia, because he claims that qualia are merely one among many categorical properties. If he is right about this, then dispositional identity will suffice for categorical identity as well.

Heil’s second way of dealing with qualia is a form of representationalism. Generally representationalism is a view on which the seemingly problematic properties of conscious experience are not properties of *experiences* but rather properties of objects represented. According to representationalists, this *exhausts* the phenomenal character of experience. Heil’s representationalism is a little different. He thinks that classic representationalism goes a long way to avoiding the apparent mystery of how there can be conscious experience in a world where there are only physical properties, but that it can’t go all the way. Heil admits to there being certain ‘residual properties’ of conscious experience that seem to escape representational characterization. He then attempts to provide a solution to this apparent problem by accounting for residual properties in a way that does not make appeal to the idea of distinct mental properties. He argues that in so far as it is reasonable to countenance non-representational content for phenomenal experience, that non-representational content can be explained by appeal to the medium of representation – where the medium of representation will be some uncontroversially physical item like a brain state or artificial sensory apparatus. I will not provide a knockdown argument for any of Heil’s proposed solutions to the apparent problem that qualia pose for physicalist ontologies like his. I will, however, argue that representationalism in particular – either of the classic sort or Heil’s unique brand – fits very

very comfortably with Heil's austere ontology. If there are only physical properties, then representationalism about qualia seems to be the most viable account of conscious experience.

I will also make clear that Heil's strategies in dealing with qualia are all *deflationary*. That is, they try dissolve the problem that qualia are meant to pose for physicalist ontologies rather than provide a solution to it. Deflationary strategies are notoriously unpopular in the literature. Many philosophers seem to struggle deeply with the idea that the subjectivity of phenomenal experience can be explained in terms of the apparent non-dispositionality of qualia, or the representational content of phenomenal experiences or by sole reference to an awareness of the medium of representation. The conviction that the subjectivity outstrips all of this runs deep. I will neither claim that Heil's deflationary strategies work against those worried about the subjectivity of experience, nor shall I say definitively that the worries about subjectivity are decisive against Heil's view. I will argue, however, that the problem of subjectivity in so far as it is construed to present an *in principle* divide between mental and physical places a dilemma in the path of non-reductive views about the mind. In so far as non-reductionists recognize that there is a problem of this magnitude posed by subjectivity, they are not really physicalist. If they think that it is solvable then there is no interesting sense in which mental properties are ontologically distinct from physical ones. One may as well accept Heil's view that there are only physical properties.

Since I will not provide positive argument for Heil's view in this paper, I hope to show at the very least that the philosophy of mind is at a crossroads. The traditional positions are no longer available. Non-reductive accounts are incoherent. Reduction is implausible and so is non-physicalism. So, either we accept something like Heil's position on which there are only physical properties, or admit that the way forward lies in some rather extreme non-physicalist views or an extreme physicalist view like reduction.

Chapter One: Setting the Scene

1.1 The Positions and Their Arguments

The thesis I will be referring to as *Reductive Physicalism* has it that mental properties just are physical properties. Adherents consider it sufficient and necessary for this sort of claim that there be a successful form of ‘reduction’. There are two important forms of reduction: *a priori* conceptual reduction and *a posteriori* theoretical reduction. Reductionists of the former sort include philosophers who think that the mental can in some sense be defined in terms of the non-mental. ‘Non-mental’ here can be taken to mean the same as ‘physically acceptable’. The supposed virtue of these views is that non-mental vocabulary plainly describes physical processes. So, behaviorists aimed to translate talk about mental states into talk about behavior. Common Sense Functionalists aim to analyze mental concepts as functional concepts. In each case, the process of reduction is carried out *a priori* and taken to establish conceptual equivalence between mental and non-mental vocabularies.² It is important that the process of defining the mental in terms of the non-mental be exhaustive. Any feature of mental phenomena that escapes analyses will be considered irreducible.

Theoretical reduction is different. It expects that we can discover, through *a posteriori* methods, laws connecting mental and physical properties. This would consist in “providing for each mental property M a necessarily co-extensive physical correlate P such that M occurs at time *t* iff P occurs at *t* as well.” (Honderich 2005) The paradigm example of a theoretical identity is the identity of water and H₂O. Here there is claimed to be no equivalence in the concepts of ‘water’ and ‘H₂O’ but nevertheless, thanks to empirical discovery, we can say that water *is* H₂O. To the extent that we can find an appropriate analogy with the case of water and H₂O, mental states could be thought of as theoretically identified with physical properties.

Arguments for *Reductive Physicalism* take the following form:

1. Reduction of either sort is possible iff mental properties just are physical properties.
2. There is a successful reductive procedure.

² This is as opposed to *a posteriori* functionalists who take the process of identifying mental concepts with functional ones to be an empirical one, revealed by science. This of course, makes it no less reductive.

3. Therefore mental properties just are physical properties.

The thesis I will refer to as *Non-Reductive Physicalism* rejects the possibility of both kinds of reduction. On this basis, non-reductionists claim that mental properties are distinct from physical ones. I'll give this claim the name, *Irreducibility*, because I will be referring back to it at various points in this paper. Note that as per my introduction, the category 'property' as it features in the formulation of *Non-Reductive Physicalism* is an ontological category. In this way, it is proper to describe such accounts as species of 'property dualism', but in a weak sense. Claims about the distinctness of mental properties, as made by the non-reductionist are not intended to falsify physicalism.³ There are a host of philosophical tools that non-reductionists have used to secure their physicalist credentials. Most important are the notions of 'supervenience' and 'token identity'. I will discuss the details of these when I assess their suitability for securing physicalism for *Non-Reductive Physicalism* in chapter three.

Arguments for *Non-Reductive Physicalism* start similarly to those for *Reductive Physicalism*. They take the following form:

1. If mental properties are physical properties, then reduction is possible.
2. There is no successful reductive procedure.
3. Therefore mental properties are not physical properties.

The position I will refer to as *Strong Property Dualism* also claims that reduction is not possible and that mental properties are distinct from physical ones. The sense in which the properties are considered distinct, however, is stronger than intended by non-reductionists. David Chalmers is a good example of a contemporary strong property dualist. I will take his view to be instructive as to the core claims of *Strong Property Dualism*. For Chalmers, mental properties – in particular the properties of conscious experience – are properties that are 'new' relative to the properties of any physical science and exist 'in addition' to them. These mental properties are also fundamental with respect to physical properties in the sense that they are not explainable in terms of them. With these new fundamental properties Chalmers takes there to be new fundamental laws as well, that link physical states to mental ones. Crucially on this view, the laws that link the mental with the physical are contingent – they happen to hold in the actual world, but had our laws been set up differently, we may have had every physical feature we actually do have, but have none of the properties of conscious experience.

³ I will go on to refer to this sort of account as Non-Reductive Physicalism; but it could have the alternative name 'Weak Property Dualism'.

This sort of position is often illuminated by an appeal to a creation myth. If our universe were as the strong property dualist says, then when God created it, He would have taken a special act of creation to ensure that there were mental properties. God would have first created all the physical substances and properties *and then* gone on to create mental ones and the laws linking them to the physical. This way of explaining *Strong Property Dualism* makes clear that mental properties are explicitly considered *additions* to the world, in contrast to *Non-Reductive Physicalism*. Indeed, this would be in contrast to the sort of creation myth that any physicalist would endorse. If physicalism were true it seems that it would have to be the case that all God would have to do in order to create all the phenomena of the actual world is to create its physical phenomena and from these the mental phenomena would necessarily follow.⁴ Non-reductive brands of physicalism would then need to proceed to complement this sort of creation myth with an account of how there can still be – in some ontologically interesting sense – a non-reductive relation between mental and physical phenomena so that they are able to distinguish their view from *Reductive Physicalism*.

David Chalmers' argument for *Strong Property Dualism* begins with the expectation that, for all physical phenomena, we can provide the sort of *a priori* conceptual reduction described above. Chalmers call it 'conceptual analysis'. Chalmers thinks, of course that conceptual analysis is not possible for some mental phenomena and so that physicalism is false as it concerns the mind. With this in mind, the argument for *Strong Property Dualism* could be put like this:

1. If physicalism as it concerns the mind is true, then conceptual analysis is possible for all mental phenomena.
2. Conceptual analysis is not possible for some mental phenomena.
3. Therefore physicalism is false as it concerns the mind.

This argument generally comes with an argument intended to support the second premise. Popular ones include the knowledge argument, the argument from the explanatory gap and the zombie argument. In so far as a further argument for premise two is required to establish *Strong Property Dualism*, only the zombie argument will be considered in this paper. I will leave the details of this argument for chapter two where I discuss the apparent relationship

⁴ One might think that this blurs the lines between *Non-Reductive Physicalism* and *Reductive Physicalism*. This may be a point worth pursuing, but I find the use of creation myths here purely illustrative and not useful in a substantive discussion about the difference between these positions.

between a commitment to zombies and a commitment to *Substance Dualism*.

The position I will refer to as *Substance Dualism* claims that not only are there distinct mental properties – along the lines of *Strong Property Dualism* – but that these properties are borne by distinct mental substances. Particular arguments for *Substance Dualism* are not important for my purposes. In so far as *Substance Dualism* features in this paper, it serves as a position that *Non-reductive Materialism* and *Strong Property Dualism* wish to avoid. I mention the arguments for *Strong Property Dualism*, *Non-Reductive Physicalism* and *Reductive Physicalism* because they will feature in my argument that there is a hidden assumption that these views all share. What arguments there are for *Substance Dualism* is not essential for this purpose.

1.2 A Hidden Premise

John Heil thinks that each of the above mentioned arguments make the following assumption:

(P): When a predicate or concept applies truly to an object, it does so in virtue of designating a [genuine] property possessed by that object and by every object to which the predicate truly applies (or would apply). (Heil and Robb 2003 p 176)

(P) is a complex principle, however and it is important to appreciate its complexity. The best way to do this is to break (P) up into three sub principles that are entailed by (P) but that do not necessarily entail (P) itself. Together, these three principles constitute (P).

The first could be set out like this.

(P₁) Predicate terms play a referential role. In particular they refer to *properties*.

There are two very good reasons why this principle is required by the above arguments: (P₁) is required to avoid circularity *and* to ensure an ontological conclusion. The premises of the above-mentioned arguments as they stand can't be claims about reality – that is, they can't be claims about mental *properties*. If they were, each argument would be circular. Instead, the premises should be considered as comments on the relationship between mental concepts or predicates. We need (P₁) to be established independently in order for these arguments not to beg the question against the opposition. On the other hand, (P₁) is also required to ensure that these arguments *have* ontological conclusions in the first place. We need to be able to move from claims about concepts and predicates to claims about the world. This requires a premise

that claims that the two are appropriately related. That premise is (P₁).

The second sub principle is:

(P₂) Predicate identity is property identity – a difference in predicate is a difference in referent.

This principle is similar to the first. The difference between them is that (P₂) is stronger. (P₁) just entails that predicates refer to something in the world, not that every predicate refers to a tailor-made property. (P₂) as much as (P₁) needs to be assumed by the above arguments. Each of these have as a premise that reduction – its success or failure – has consequences for the world. As per my argument immediately above, we ought to conceive of reduction as first and foremost a relation between predicates and/or concepts. If unique predicates did not pick out unique aspects of the worlds, there would be no ontological consequence for either showing that these concepts are co-extensive with or analyzable in term of physical ones. Neither would there be any ontological consequence for showing that this is not possible. In other words, if it were not for presupposing (P₂) we'd have no reason to have faith in the ability of conceptual analysis or theoretical identity to secure mental-physical property identities. And nor would we have faith in their failure to show mental and physical properties are distinct.

This particular principle seems to trouble Heil quite a lot. He says that it gives us the ability to 'read off' the nature of reality from the semantic properties of predicates. So, if there are irreducibly *mental* predicates, there are irreducibly *mental* properties as well. (P₂) allows us to investigate the nature of reality from the armchair.

The third sub-principle is:

(P₃) When the same predicate is applied to different things, it is because of a common referent.

This is essentially the claim that the referents of generally applicable predicates are 'universals'.⁵ Universals are peculiar things. They are considered as 'wholly present' in the items they characterize.⁶ Universals can be numerically identical, even if the objects they

⁵ Whether this sort of view about universals is plausible in its own right is not important here as yet. What is important is that without such an ontology of properties, universalism would not be able to provide an ontology that compliments (P), nor the arguments that it assumes.

⁶In this paper, I will make the '*in re*' assumption about universals.

characterize are in different spatial locations. Universals are helpful when it comes to generally applicable predicates precisely because they seem to account for how it is that the same predicate can be true of distinct items. That viewing the referents of mental predicates as universals is an assumption of the arguments for traditional views is not as easy to see. We might see it as quite straightforwardly following from (P₁) and (P₂): If predicates refer to properties, and identical predicates refer to identical properties, then in cases where the same predicate applies to distinct items, there will be something that the two share in virtue of which the predicate is applied to them.

Consider as well premise two of the argument for *Non-Reductive Physicalism*, which claims that reduction is not possible in either form. An important argument for the truth of that premise comes from an appeal to the multiple realization of the mental. In this paper, I will define ‘multiple realizability’ like this:

Multiple Realizability: A single mental predicate *m* can truly apply to diverse physical systems.

Note that this is not a traditional formulation of *Multiple Realizability*. Usually it is defined in terms of the possession of identical mental *properties* by diverse physical systems. In the context of this paper (indeed in others as well) I think this way of characterizing *Multiple Realizability* is ill advised. Articulating *Multiple Realizability* in terms of mental properties is in danger of begging the question. With this formulation of *Multiple Realizability* in hand, an argument for *Non-Reductive Physicalism* that appeals to it might go like this:

1. If a single mental predicate *m* can truly apply to diverse physical systems, then there is no physical predicate with which the mental predicate is identical.
2. A single mental predicate *m* can truly apply to diverse physical systems.
3. If there is no physical predicate with which the mental predicate is identical, then, given 2, mental properties are distinct from physical ones.
4. Therefore, some mental properties are distinct from physical ones. (Heil 2001 p 7)

Consider premise two of this argument. If it were not the case that the referent of the mental predicate were numerically identical between spatially disparate items, wholly present in each of those items, I don’t think that we would worry that the referents of mental and physical predicates are distinct. This argument could also be used as one for *Strong Property Dualism*

in so far as *Multiple Realizability* is also a consideration against conceptual analysis. For similar reasons, an ontology of universals also seems to suit the robust ontology of mental properties espoused by the strong property dualist. Consider premise two of the argument from multiple realizability fleshed out like this:

- a. There exists v, w, x, y and z.
- b. v is James,
- c. w is Voltron, the Martian
- d. x are James' physical aspects
- e. y are Voltron's physical aspects
- f. z is the referent of 'is in pain.'
- g. z is wholly present in x and y.
- h. z = z.
- i. 'is in pain' refers to no physical aspect of James nor Voltron.

Whatever the sorts of things mental predicates refer to on *Strong Property Dualism*, it would make sense to say that they are universals.

1.3 (P)'s Restrictions on Philosophy

Heil argues that (P) places the following restriction on philosophical theorizing:

- (R) Where *Gs* are presumed to be uncontroversial items—those posited by the physical sciences, for instance—and *Fs* are putatively higher-level items; if talk of *Fs* cannot be analysed, paraphrased, wholly decomposed into talk of *Gs*, either *Fs* are distinct from *Gs* or there are no *Fs*. (Heil 2003 p 51)

If (P) is true then the failure of reduction has *ontological* consequences.⁷ In a world where predicate identity is property identity it seems natural to pursue reduction, admit distinct mental properties, or face eliminativism. Either there is nothing that the irreducible predicates refers to, in which case we have eliminativism. Or irreducible predicates have their own unique referents – irreducible properties – in which case we need to choose between *Non-Reductive Physicalism* or *Strong Property Dualism*. Since eliminativism is so unattractive in

⁷ Recall that 'property' is being used here as an ontological category in the sense that whatever sort of thing a property is, it is a feature of mind independent reality.

philosophy of mind, philosophers have been motivated to pursue either of these views or *Reductive Physicalism*. (Heil 2003 p. 24) Given a bias towards physicalism and against reduction, however, we seem left only with *Non-Reductive Physicalism* as an account that in so far as it rejects reduction, but admits to irreducible higher-level properties. To the extent that this view ostensibly rejects reduction while preserving physicalism, this view has received wide acceptance in the philosophy of mind. The view, however, has been challenged by some serious difficulties.

Chapter 2: The Traditional Problem with Non-Reductive Physicalism: Mental Causation

The most notorious problem facing the non-reductionists is accounting for the causal efficacy of higher-level mental properties. I will argue that the non-reductionist faces a dilemma in the context of accounting for mental causation: Either he accepts that mental properties have causal powers over and above the powers that physical properties have – in which case he has falsified physicalism – or he admits that mental and physical properties have identical causal powers – in which case his view is no longer non-reductive in any interesting sense. My argument in support of this dilemma will consist broadly of a defense and critical discussion of Jaegwon Kim's critical literature on *Non-Reductive Physicalism* and mental causation.

Before I get to Kim's argument, I need to set out two metaphysical principles assumed by non-reductionists and that feature prominently in Kim's argument. The first, I'll call *Closure* and I'll define it like this:

Closure: Every event has a physical antecedent cause sufficient for the instantiation of a given effect.

This principle needs to be considered among the core claims of *Non-Reductive Physicalism* in so far as that position aspires to be *physicalist*. Another important commitment of *Non-Reductive Physicalism* is 'supervenience'. Again this seems to be a requirement of a minimal physicalism. If the non-reductionist is going to reject reduction, he will need to tell an appropriate story about how mental properties depend on physical ones such that the former are still considered physically acceptable. This is generally considered to non-reductionists to be achieved by positing a supervenience relation between mental and physical. I will discuss the supervenience relation in more detail in the next chapter. For the purpose of this chapter, however, the following rudimentary definition of supervenience will be sufficient:

Supervenience: "Mental properties supervene on physical properties in the sense that if something instantiates any mental property *M* at time *t*, there is a physical base property, *P* such that the thing that has *P* at *t*, and necessarily anything with *P* at a time, has *M* at that time. (Kim 1998 p. 39)

The reason I take a relatively uncritical stance on the supervenience relation itself in this

chapter is because it is *assumed* in Kim's argument that *Non-Reductive Physicalism* has a genuine problem with mental causation.

Another important commitment of the non-reductionist is realism about mental properties. This is not a commitment that is demanded by *Non-Reductive Physicalism's* commitment to physicalism, but rather from a general requirement that adequate accounts of mental phenomena must avoid eliminativism.⁸ What it is for something to be real is certainly a vexed philosophical question. I like what J.L Mackie said on the matter and I think it gives us an intuitively plausible starting point. He said that to be 'real' is to be "part of the fabric of this world." (Mackie 1977 p15) What might it mean then, for mental properties to be part of the fabric of the world? Whatever the very lengthy answer to this question is, I think one could at least agree that they must have the ability to make a causal difference to it. (Kim 1990 p. 202) It seems odd at the very least if one were to affirm the reality of something without at the same time affirming its ability to make a causal difference to the reality of which it is meant to be a part.

There is, however, more to be said about exactly what the causal efficacy of real mental properties consists in – at least there is if we are to understand Kim's criticism of non-reductive accounts of mental causation. When it comes to the causal efficacy of the mental, Kim seems to expect two different standards from reductionist and non-reductionist accounts of mind. For reductionists, the ways in which mental properties are causally efficacious are just the ways physical properties are since the former are reducible to the latter. For the non-reductive materialist he seems to expect more – he claims that on their view, in order for mental properties to be causally efficacious they must: "bring with them new causal powers, powers that no underlying physical biological properties can deliver." (Kim 1993 p. 204) This might be considered a controversial claim, and I will comment on its plausibility later. For now, just notice that the textual evidence indicates that it is at least *Kim's* conception of the realist commitments of non-reductive materialists; and it seems to be the one he had in mind when he has criticized non-reductive accounts of mental causation.⁹

⁸ I will not put forward argument for why eliminativism is a bad thing, I will merely assume that it is.

⁹ In *Mind in a Physical World* Kim asks whether it is the redness of a bullfighter's cape, or its secondary property of 'provocativeness' that *causes* the bull's anger. After arriving at the conclusion that the cape's redness was sufficient, he expresses puzzlement over what "*further* work causal work is left for its 'provocativeness' to do? What *special* contribution of its own can the cape's provocativeness make in the causation of the [bull's] anger?" (Kim 1998 p. 53) I recognize that there is not a complete analogy between this sort of case and the case of higher order mental properties – there is no necessary supervenience relation between the cape's redness and its provocativeness. But I do think this passage reveals what Kim thought to expect in terms of causal efficacy form irreducible properties.

The problem of explanatory exclusion arises against this backdrop of metaphysical assumptions. With them in hand, Kim believes he can construct a case against non-reductive materialism on the basis of its inability to account for mental causation. His case begins with the *Principle of Explanatory Exclusion*, which states that:

Principle of Explanatory Exclusion: “Two or more complete and independent [causal] explanations of the same event or phenomena cannot co-exist.” (Kim 1993 p. 250)

We now have the apparatus in place to appreciate how Kim sees the problem of mental causation arising for the non-reductionist. Consider the following two events:

M: ‘my desire for fresh air’

P: ‘my rising and going for a walk.’

Consider further a scenario in which they are in a causal relation to one another – my desire for fresh air causes my rising and going for a walk. That is, *M* causes *P*. In this scenario, we have a mental event *m* that occurs at a particular time, *t* that allegedly causes a particular physical event, *P*. However, remaining true to *Closure*, the physicalist is obligated to claim as well that *P* has some sufficient physical cause, *P** at *t*. (Kim 1998 p. 37) The trouble is that this seems to call into question the causal efficacy of *M*. “Given that *p* has a [sufficient] physical cause *P**... what causal work is there left for *M* to contribute? The physical cause here seems to be threatening to exclude the mental one.” (Kim 1998 p 37) And if we admit that *m* does some causal work here, we seem to be committed to overdetermination. The non-reductionist needs to give an account of the causal efficacy of *M* that does not violate *Closure* or any of the core claims of *Non-Reductive Materialism*. I’ll start with the non-starters and then move on to the more to the serious options.

The first option, Kim calls the ‘partial cause option’ and it envisages that each *M* and *P** are neither sufficient but each necessary parts in the total sufficient cause of *p*. Here we have the benefit of avoiding overdetermination since *M* and *P** are not independent and complete causes of *p*. However, here we *do* seem to be violating *Closure* as we are allowing that a physical event does not have a sufficient antecedent physical cause. The second option he calls the ‘overdetermination option’ and it envisages *M* and *P** each being independent and sufficient causes for *P*. This, however, doesn’t seem to be what any physicalist thinks is going on when the mental is exercising its causal powers – namely that a mental event on its own can form a sufficient cause for a physical event. The third option is the ‘causal chain option’, which envisages that “two conditions can each be a sufficient cause of some single event by

being different links in the same causal chain leading to the effect event.” In this case, the later cause is causally dependent on the earlier one.” This is not a viable option since it requires *M* and *P** to occur at different times when it is very possible that they could occur simultaneously. This option also seems to violate *Closure* in that the physical event *P** is not envisaged as a sufficient cause of *P*. (Kim 1990 p. 40 and Kim 1993 p. 250)

Kim thinks that the only viable options that remain are the ‘identity option’ and the ‘supervenience option.’ With the identity option we have the benefit of there being only one causal story to tell – mental causes are identical to physical causes. This option, however, is clearly at odds with *Non-Reductive Physicalism*. (Kim 1990 p 42) So, the only serious option for the non-reductionist Kim thinks is left over, and the one I will consider in this paper, is the supervenience option, which appeals to *Supervenience* as a means of securing the causal efficacy of mental properties. This appeal to *Supervenience* gives rise to the phenomenon of ‘supervenient causation.’ As with *Supervenience* it comes in reductive and non-reductive varieties; but we can put that aside for now and concentrate for the moment on how the concept of *Supervenience* can be made use of in securing the causal efficacy of higher order mental properties. I will base my discussion of supervenient causation on Kim’s discussion of it in his *Epiphenomenal and Supervenient Causation*. (Kim 1993a)

There, Kim begins by asking his reader to consider the notion of ‘epiphenomenal causation.’ To illustrate, consider successive reflections of a dancer in a mirror. The mirror reflects one state of the dancer, *D*, at time *t* and then another immediately after it, *D**, at time *t*₂. Here, the succession of reflections *D* and *D** can be accounted for fully by appealing to causal processes occurring at a more fundamental level – in this case it would be the dancer herself and her movements. This is a feature shared by all epiphenomenal causal processes. (Kim 1993 p 93 & 103) However, Kim points out that calling a causal process ‘epiphenomenal’ is not to suggest that events that play a part in that process are always *epiphenomena* in the sense that they are causally inert. (Kim 1993 p. 94) This is important because Kim wants to say that causal processes involving mental events are cases of epiphenomenal causation, without saying that mental events are epiphenomena. However, it surely is the case that with epiphenomenal causal processes like the one involving successive mirror reflections, the events appealed to *are* epiphenomena. Kim then needs to distinguish between cases of epiphenomenal causation like that of the successive mirror images from those involving mental epiphenomenal causation such that the latter, while being epiphenomenal in the sense that they can be accounted for by causal processes that occur at a more fundamental level, do not render mental events causally inert. (Kim 1993 p. 94)

To achieve this distinction, Kim formulates an attenuated variety of epiphenomenal causation, which he dubs ‘supervenient epiphenomenal causation.’ (Kim 1993 p. 94) By invoking *Supervenience* Kim suggests that it’s possible to secure a form of non-redundant causal efficacy for the mental. He says a case of supervenient causation takes place between two events just in case the following scenario holds:

“There exist two events [x and y] characterized by certain properties such that, x has F and y has G, where F and G are higher level properties and x’s having F supervenes on x’s having m(F) and y’s having G supervenes on y’s having m(G), where m(F) and m(G) are lower level properties on which F and G supervene respectively. Further there needs to be an appropriate causal connection between m(F) and m(G).” (Kim 1998 p 39)

What this pattern shows is that higher level ‘apparent’ causal relations *themselves* can be taken to be supervenient upon an underlying causal process by virtue of the supervenience of higher level properties / events that constitute the higher level causal process supervening on properties / events that characterized the genuine causal process at a lower level. (Kim 1993 p. 105) In short, supervenience allows for mental causation like this:

“Mental event *M* causes a physical event *P*, because *M* is supervenient upon a physical event *P** and *P** causes *P*. Similarly, when mental event *M* causes another mental event *M**, this is so because *M* supervenes on [some] physical state and *M** on [some other physical state]” (Kim 1993a p. 106)

To see how the concept of supervenience might secure causal efficacy for the mental, let’s apply the theory of supervenient causation to the above-mentioned case of mental to physical causation – my ‘desire for fresh air’ causing my ‘rising and taking a walk.’ There is clearly a purely physiological account of the causal process occurring in my body here, Kim assumes, and I won’t argue with him. He says: “If a belief desire pair, *M* causes, bodily movement, *B*, it is highly plausible for a physicalist to suppose that the total physical state of the agent at the time provides a full cause of *B*.” (Kim 1991 p. 56) Furthermore, Kim thinks it can’t be that my desire *qua desire* acts directly on my muscles involved in my rising and walking, causing them to contract and so on.¹⁰ Rather, if my desire is to play a causal role here, it must in some way ‘make use’ of the underlying physiological causal process that is clearly at work. (Kim 1993a p 103) It is this ‘*making use of*’ lower order causal processes that I think is meant to give causal processes involving higher order supervening properties what cases like the successive mirror images don’t have – a necessary dependence relationship on the lower level

¹⁰ And he’s surely right about this.

causal processes that ultimately account for them. Importantly, Kim notes that this approach does not award *the same sort* of causal potency to mental events as it does to fundamental physical processes. But it does stop short of leaving mental phenomena causally inert. (Kim 1993a p 106)

Now, let's infuse this scenario with the metaphysics of reduction and non-reduction. Kim envisages supervenient causation to allow for the reduction of higher order causal processes to lower level causal processes. His model of the reduction of a higher order causal process to lower order ones mirrors the relation he envisages between higher and lower order mental properties in cases of supervenient causation. He writes:

“If the [higher order] causal relation to be reduced is one from an instance of property F to an instance of property G, we need to correlate F with some [lower level] property m(F), and also G with m(G), and then show that m(F) and m(G) are appropriately causally connected.” (Kim 1993 p 97)

It is no mystery *here* then, why *Supervenience* can work for the good of mental causation. On this picture, the causal powers of the supervening higher order mental property are just the causal powers of the physical base on which it supervenes, then mental properties affect the world in so far as the physical properties on which they supervene affect the world. In this spirit, Kim constructs the *Principle of Causal Inheritance*, which he articulates like this:

Principle of Causal Inheritance: “If *M* is instantiated on a given occasion by being realised by *P* [or by *P* as its supervenience base being instantiated], then the causal powers of *this instance* of *M* are identical with (perhaps a subset of) the causal powers of *P*.” (Kim 1993 p. 208)

Kim will not, however, want to say the same for a non-reductive construal of supervenient causation. Considering *Irreducibility*, Kim will expect the non-reductive materialist to say that the causal processes at higher levels involving higher level processes are not reducible to lower level causal processes – that the causal processes involving higher level properties are ‘something over and above’ the causal processes that occur at the level of the supervenience base. (Kim 1993 p192) In this way the non-reductive materialist cannot accept the *Principle of Causal Inheritance*. (Kim 1993a p 209) As we saw earlier, Kim expects that given *Irreducibility* mental properties have causal powers over and above the causal powers of the physical bases on which they supervene. This is a crucial assumption of Kim's criticism of non-reductive accounts of mental causation as it features overtly in his ‘supervenience

argument' against these accounts. It has also proven in discussion to be a controversial understanding of the realist commitments of non-reductive materialism. So, before I get to the supervenience argument, I will pause for a moment and consider whether this is a fair interpretation of the realist commitments of non-reductive materialism. My objective here is just to show that Kim's understanding of the realist commitments of non-reductive accounts of mental causation is acceptable on the face of it.

Let's reconsider *Irreducibility*. This is what John Searle has said about it: he writes that to consider that the mental is not reducible to the physical means that the mental is "...something that exists in its own right...mental states are something *distinct from and over and above* neuro-biological states." (Searle 2002 p 58)¹¹ And again later "The property dualist means that *in addition* to all the neurobiological features of the brain, there is an extra, distinct, nonphysical feature of the brain." (Searle 2002 p. 61) So, at least we know that Kim is not alone in his interpretation of *Irreducibility*. Now, it seems to me, at least on the face of it if we are going to consider properties so characterized as real properties that make a causal difference to the world of which they are apart, they are going to have to do so *as distinct properties*. This means that they cannot only have causal efficacy in virtue of whatever physical features they have. For irreducible properties to be real properties, it seems quite fair to expect from them to make a causal difference to the reality of which they are a part in ways that are distinct from physical properties.

One might contest this conception of the realist commitments of non-reductive materialism, saying that perhaps we are asking too much from the non-reductive materialist and that mental properties do not need to have causal powers *over and above* physical properties. However, I think that such a response would need to be followed by a good explanation as to how we can carry on calling mental and physical properties *distinct* from physical properties. I am with Kim in thinking that: "...unless mentality made causal contributions that are genuinely novel, the claim that it is a distinct and irreducible phenomenon over and beyond physical-biological phenomena would be hollow and empty." (Kim 1993 p. 204)¹² Here, even an appeal to *Supervenience* isn't of much help to the non-reductive materialist. As long as the casual powers of mental properties are wholly derivative from those of physical properties – even if we are to consider them as supervenient on those physical properties – reduction as

¹² He expresses this sort of sentiment elsewhere when he writes: "The implications of the Causal Inheritance principle are devastating for non-reductive physicalist: if the causal powers of M are identical with thus of its relation [or supervenience] base, then M in effect contributes nothing new causally and M's claim to be an irreducible property is put in jeopardy." (Kim 1993a p 209)

opposed to *Irreducibility* seems to be a more attractive understanding of the relationship between mental properties and the physical properties on which they supervene.

From the preceding I hope to have shown that there is at least prima facie plausibility that given *Irreducibility*, something like novel causal powers are to be expected from mental properties. I will now turn my attention to Kim's 'supervenience argument.'

In the supervenience argument, Kim makes the case that if the non-reductive materialist is to stick with the theses of *Irreducibility*, *Closure* and *Supervenience* – which it seems he must – he'll ultimately render mental properties causally inefficacious. The argument has undergone some development over the years. Kim formulates it first in his *Mind in a Physical World* (1998) and then again in *Physicalism, or Something near Enough* (2005). I will reflect this development in my exposition of the argument. I will use his earlier formulation for the general structure of the argument and appeal to the later version where it is clear that the argument encounters difficulties. Kim formulates his argument in the form of a dilemma and it goes something, like this:

Supervenience Argument:

- (1) Either *Supervenience* holds or it fails.
- (2) If *Supervenience* fails, given *Closure* and *Irreducibility*, there is no visible way of understanding the possibility of mental causation.
- (3) If *Supervenience* holds, given *Closure* and *Irreducibility*, there is no visible way to understand the possibility of mental causation.
- (4) Therefore, if we accept *Supervenience*, *Closure* and *Irreducibility*, there is no visible way of understanding the possibility of mental causation. (Kim 1998 p 39; Crisp and Warfield 2001 p306)

The purpose of (2) is to illustrate the indispensability of *Supervenience* to non-reductive materialism. This is not an uncommon conclusion to draw from this premise; indeed many non-reductive materialists have done so. (Fodor 1987 p. 42) This is because, as Kim notes, "...supervenience brings mental phenomena within the ambit of the physical: the physical determines the mental, and in that sense, the mental does not constitute an ontologically independent domain intruding from the outside." Kim thinks that the failure of supervenience, given *Irreducibility*, would result in the mental domain "floating free" of the physical domain. (Kim 1998 p40-41) In this way, *Supervenience* cannot fail if we are to have an intelligible

non-reductive account of mental causation. I will take this as a relatively uncontroversial horn of the dilemma and spend my time in explanation and defense of (3).

Kim's defense of (3) begins with his asking his reader to consider, given *Supervenience* and *Irreducibility*, an instance of mental to mental causation. Namely:

M causes M* (M-M*): An instance of a mental property *M* (fear of the dark) causing another mental property *M** (anxiety upon entering a darkened room). (Kim 1998 p 41)

It is clear that if *Supervenience* is going to hold in a situation like this, our consequent mental state *M** is going to have a physical base on which it necessarily supervenes, call it *P**. What are we to say, however, in response to the question of how *M** became instantiated? There are two possible answers according to Kim, between which there exists an alleged tension: either we can account for the instantiation of *M** by citing *M* as a cause, or we could cite the instantiation of *P**. (Kim 1998 p. 42) Considering *Supervenience* holds of **M-M***, we are obligated to say that *M** necessarily owes its instantiation to the instantiation of its supervenience base, *P** – the instantiation of *P** is a sufficient condition for the instantiation of *M**. Crucially, Kim thinks that this means that *P** will instantiate *M** no matter the mental event that preceded it – in this case no matter whether *M* did; “as long as *P** is instantiated, *M** must occur; it need not have been *M* that did.” (Kim 1998 p. 42) In saying this, Kim seems to be affirming the counterfactual “if it were not for *M*, *P* would still have been able to act as the sufficient cause of *P**.” This is a controversial step in the argument, but it is not appropriate for me to deal with it now. For the time being I will name this feature of Kim's argument the *Exclusionary Counterfactual*, so we can keep track of it when it rears its head again later. For now, however, notice that all this considered, at least some explaining needs to be done if we are going to secure for *M* a causal role in the causal chain culminating in *M**.

The only solution Kim thinks is on offer to someone trying to salvage the causal efficacy of *M* is to consider it as the cause of *M**'s supervenience base, *P**. From this Kim draws the general principle that: “To cause a supervenient property to be instantiated, you must cause its supervenience base to be instantiated.” (Kim 1998 p. 42) And this seems quite acceptable on the face of it. Indeed it seems a common practice for those motivated to alter higher order supervenient properties. Chemists produce medicines that alter the neural structure of the body to relieve pain; artists who wish to add to the beauty of their work will do so by altering its physical properties. In this way, Kim hopes to show that **M-M*** implies the possibility of mental to physical causation. (Kim 1998 p. 43) If the non-reductive materialist is to salvage the causal efficacy of *M*, according to Kim, he would then need to accept the following:

M causes P* (M-P*): *M* (fear of the dark) causes *M** (anxiety upon entering a darkened room) by causing *P** (the supervenient base property on which *M** supervenes.)

This is what Kim calls ‘downward causation’ and it allegedly just replaces one puzzle with another. Since *Supervenience* is going to hold in this situation as well, how might we make sense of *M* as the cause of *P** if, as we cannot deny, *M* too has a supervenience base, *P*, that could be considered the sufficient cause of *P**? (Kim 1998 p. 42) If we are to hold on to the idea that *M* causes *P**, we are committed to both *M* and *P* being causes of *P**.” (Kim 2005 p. 41) It is at this point that the problem becomes one for the non-reductive materialist in particular. According to Kim on *Irreducibility*, *M* is distinct from – something over and above – *P* and thus its causal powers are also thus something over and above the causal powers of *P*. (Kim 2005 p. 42) In this way Kim envisages the non-reductive materialist offering two distinct causes for *P** that occur at the same time.¹³ About this sort of situation in his earlier formulation of the supervenience argument, he says that “...both *M* and *P* seem severally eligible as a sufficient cause of *P**.” (Kim 1998 p. 44) This would mean that *P** is causally overdetermined and by extension so would every instance of mental to physical causation. This, however, is surely wrong. Since the non-reductive materialist will be affirming *Supervenience* while positing both *M* and *P* as a cause of *P**, he cannot, and is not, saying that *M* on its own can serve as a sufficient cause of *P**. Since *M* is a supervenient property, there is no such thing as *M* on its own. This is important because it seems that it allows the non-reductive materialist to include *M* and *P* as a cause of *P** without infringing the *Principle of Explanatory Exclusion*.

This is certainly an oversight on Kim’s behalf, but perhaps we can put it down to the argument being in its formation stage. In his later formulation (2005) Kim does away with the *Principle of Explanatory Exclusion* as formulated above. He replaces it with another principle, which is put to work for the same exclusionary purpose as the *Principle of Explanatory Exclusion* was previously. He calls it the *Causal Exclusion Principle* and he formulates it like this: “If an event *e* has a sufficient cause *c* at *t*, no event distinct from *c* can be a cause of *e* (unless this is a genuine case of overdetermination)¹⁴. (Kim 2005 p17) Here

¹³ This is because *M* and *P* are in a relationship of supervenience; and supervenience base properties are always instantiated simultaneously with their supervenience bases.

¹⁴ A case of genuine overdetermination is one in which there are two independent and sufficient causes for the same even. Consider the case of a man being shot in the heart with two bullets simultaneously. We would not fall foul of the *Principle of Causal Exclusion* if we were to say that each bullet was a sufficient and independent cause of the man’s death. However, since these are not the situations we envisage when mental causation is at issue, non-

Kim seems to be tailoring his argument to fit what the non-reductive materialist is trying to say with his thesis of *Irreducibility* – that although *M* and *P* are not complete and independent when envisaged as the causes of *P**, they certainly are distinct. *This* is certainly a principle which non-reductive brands of supervenient causation violate. (Kim 2005 p. 42) With this new principle in place, Kim hopes that the shortcomings of non-reductive materialism may be exposed more noticeably. This is what they seem to me to be, bearing this new principle in mind: ‘*P* is sufficient for *P** which in turn is sufficient for *M**; *M*, while not sufficient for *P**, is distinct from *P* and also causes *P**’. But what role could *M* be playing here if we already have a sufficient cause for *M**? It doesn’t seem available to the non-reductive materialist to reject *P*’s sufficiency for *P** nor *P**’s sufficiency for *M**. The former would violate *Closure* and the latter would violate *Supervenience*.

This may not lead to overdetermination in the sense that we have two independent sufficient causes explaining the same event, but we certainly do seem to have an overabundance of causes of some sorts¹⁵. This is clearly not an instance of traditional overdetermination since the two causes are dependent on each other – let’s call it ‘awkward overdetermination’. Its awkwardness stems from *M*’s being considered distinct from *P* while necessarily depending on *P*. In this way, its awkwardness is due to the metaphysics of non-reductive materialism.

Kim concludes that the following sort of picture follows if we accept non-reductive materialism: *P* causes *P** and *M* and *M** supervene on each respectively. There is only one causal process going on here and it’s between *P* and *P**. There is no more causality between *M* and *M** here than there is between the successive shadows of cars – all the causal work that needs to be done is done by *P*, *M* is superfluous. Here we have lost the causal efficacy of the mental. (Kim 2005 p. 21-22) The correlations between instances of *M* and *M** are not accidental, according to Kim. They are “law-like and counterfactual sustaining regularities arising out of *M* and *M**’s supervenience on *P* and *P**.” These correlations, however, only give us an appearance of causation. For Kim it is crucial to distinguish “between genuine, productive and generative causal processes on the one hand, and the non-causal regularities that are observed because they are parasitic on real causal processes.” (Kim 1998 p. 45)

With this in mind let’s turn back to the *Exclusionary Counterfactual*. Kim expresses similar sentiments in other work. In *The Non-Reductivist’s Troubles with Mental Causation* (1993b)

reductive materialists cannot avoid this principle by appealing to mental causation bang a genuine case of overdetermination.

¹⁵ Kim expresses something like this sentiment when he writes: “However, our problem is not exactly that of causal determination, but [it does] have to do with an overabundance of causes. The difficulty is that the causal status of the dependent event is threatened by the event on which it depends.” (Kim 1998 p53)

he writes: "...if it is the case [that] *M* has no causal work to do on its own, why not *abolish it*?" (Kim 1993b p 209)¹⁶ This sort of commitment to the explanatory redundancy of *M* given non-reductive materialism seems to commit Kim to the acceptance of the *Exclusionary Counterfactual given non-reductive materialism*. The trouble for this sort of counterfactual, however, is that a commitment to *Supervenience* seems to render it impermissible as a premise in support of Kim's conclusion against a non-reductive account of mental causation. This is because, assuming *Supervenience* we cannot alter *M* without altering *P*. In effect, this is what Kim is asking us to do when he recommends that we could "abolish *M*" or that "*P** will occur no matter if *M* does." However, since it is impossible to alter *M* without altering *P*, we seem to remove *P* as the sufficient cause for *P** if we do away with *M*. In this way, mental properties seem to become causally relevant.

This is a tricky problem, but I think Kim can deal with it both directly and indirectly. He can deal with it directly by showing the counterfactual does not secure causal efficacy for *M* in the right way – that is, casual efficacy of *M* as *distinct from P*. Let's assume that Kim is right and *M* is reducible to *P*. Here we still would not be able to affirm the counterfactual "If not *M*, then *P* would still instantiate *P**" but it will be because *M* just is *P*. In this way, the counterfactual that is meant to secure the causal relevance of *M* is silent on the metaphysical status of *M* and its relationship to *P* and is subsequently not much help to any *non-reductionist* ambitions to secure the causal relevance of *M*. (Kim 1998 p. 70)

The indirect route would be to emphasize what I have called 'awkward overdetermination' as a consequence of a non-reductive account of mental causation. Recall that this is a result of claiming that both *M* and *P* are causes of *P** but are in some way distinct from each other. Searle seems to be in agreement with Kim as to why this sort of scenario is problematic. He writes:

"...we know, for example, that when I raise my arm there is a story to be told at the level of neuron firings, neurotransmitters and muscle contractions that is sufficient to account for the movement of my arm. So, if we are to suppose that consciousness also functions in the movement of my arm, then it looks like we have two *distinct* causal stories, neither reducible to each other...my bodily moments have too many causes." (Searle 2002 p. 59)¹⁷

The fact that there are "too many causes" here does not mean that the two causes need to be considered as independent from each other. *Irreducibility* – even in concert with

¹⁶ My emphasis.

¹⁷ My emphasis.

Supervenience – gives us all we need to be suspicious of the causal role *M* is meant to play here as a property distinct from *P* in the instantiation of *P**. (Kim 2005 p. 48) If the causal powers of *M* were distinct from those of *P* in that the former is merely a ‘re-description’ of the causal powers of the latter, then the situation would not be as problematic. (Kim 1998 p. 49 and 2005 p 48) But *Irreducibility* seems to be committed to a form of distinctness more substantive than this. One can’t help but feel that *M* is left dangling *in some sense* in the causal process from *M* to *P** given *Irreducibility*.

Therefore, in so far as one accepts higher-level mental properties, one seems to take on a significant burden in the form of articulating how it is that those properties are causally efficacious. I hope to have shown in this chapter that mental causation is problematic for the non-reductionist in the sense that he either has to countenance a view regarding the nature of mental properties on which they have causal powers distinct to those of underlying physical ones – in which case he is no longer a physicalist – or he has to identify the causal powers of mental properties with the causal powers of physical properties – in which case it is no longer clear in what sense his is a *non-reductionist* view. In this way, the problem of mental causation on its own threatens the availability of *Non-Reductive Physicalism* as a viable account of mind. In the next chapter I will consider reasons independent of those given in this chapter about what threatens the availability of *Non-Reductive Physicalism*.

Chapter Three: Redefining The Landscape

In Chapter one, I presented four apparently distinct accounts of mind. In this chapter, I will argue that viewing the philosophical landscape this way is mistaken. My core aim in this regard will be to show that *Non-Reductive Physicalism* is incoherent in the sense that however we characterize it, it either isn't a *non-reductionist* account or it is not a *physicalist* account. En route to defending this claim I hope to show as well that there is no substantive difference between *Strong Property Dualism* and *Substance Dualism*. In so far as one accepts *Strong Property Dualism*, one is either committed to *Substance Dualism* or cannot avoid it. If the arguments of this chapter go through, it will make being limited to the traditional positions all the more troubling. If *Non-Reductive Physicalism* is unavailable, it seems that in so far as we value physicalism, we need to persevere with reduction and that in so far as we value non-reduction, we might have to admit a rather extreme version of non-physicalism.

My strategy will be to first ask how *Non-Reductive Physicalism* can distinguish itself from *Strong Property Dualism*. I will consider three ways in which a non-reductionist could achieve this distinction. The first is to claim that, *Strong Property Dualism* is not able to avoid dualism about substances, but that *Non-Reductive Physicalism* is. The second is to argue that the relationship between *Strong Property Dualism* and *Substance Dualism* is even stronger – that *Strong Property Dualism* is committed to *Substance Dualism* – but that *Non-Reductive Physicalism* is able to avoid the same commitment. While a welcome result for the non-reductionist, if either of these lines of argument is successful, it will be a problem for the strong property dualist in so far as he sees substance physicalism as a virtue of his view.¹⁸ If he cannot rule out *Substance Dualism* then there will be nothing in the core claims of *Strong Property Dualism* that makes dualism of substance any less plausible than *Strong Property Dualism* itself. And if *Strong Property Dualism* entails *Substance Dualism* then, of course *Strong Property Dualism* is not really a version of property dualism at all, but a dualism of substance as well. The third way to distinguish *Non-Reductive Physicalism* and *Strong Property Dualism* is by appeal to supervenience. Generally this is done by pointing out that the former view has a version of the relation that tethers the mental to the physical in a tighter sense than the latter.

With regard to the first two ways, I will argue that whether it is the accusation is that *Strong Property Dualism* is consistent with *Substance Dualism* or that it entails *Substance Dualism*

¹⁸ This is generally the case.

that is at issue, the problem will be the strong property dualist's commitment to so-called 'zombie worlds'. These are worlds that are identical to the actual world in every physical / and functional aspect, but differ in that they lack conscious experience. In these worlds, exist 'philosophical zombies', creatures that are molecule for molecule identical with humans and display identical functional properties as well, but lack conscious experience.

Generally the possibility of zombie worlds is meant to imply only that mental and physical properties are distinct, while preserving the idea that there is only physical substance. I will argue that the possibility of zombies is consistent with *Substance Dualism* due to claims about the purpose and truth conditions of modal claims in debates about the ontology of mind. I will argue that modal claims – in so far as they are useful in illuminating the nature of reality – need to be viewed as made true by features of the actual world. If we view modal claims this way, then one can argue, I think, that among the possible features of the actual world that make true the possibility of zombies, is that the actual world is one in which mental properties are not borne by physical substances. With regard to the entailment of *Substance Dualism* by *Strong Property Dualism*, I will argue that such an entailment exists because the commitment to zombies entails a commitment to 'ghosts', where ghosts are conceived of as entities with exhaustively mental natures. This is problematic for the strong property dualist in so far as the very consideration that allow the possibility of zombies will in effect allow a Cartesian style argument for dualism.

Given how problematic zombies are, the most obvious way for the non-reductionist to insulate himself from the strong property dualist's relationship to *Substance Dualism* will be to show that on his view, zombies are not possible. The best way to do this is to appeal to a supervenience relation on which there cannot be physical identity and mental difference. I will present the task ahead of the non-reductionist in this regard as one of defining the modal strength of his supervenience relation such that the relation between the mental and the physical is tighter than it is on *Strong Property Dualism*. I will argue that there is no other notion of necessity available to the non-reductionists to which he could appeal and still remain true to his core claims. The only available modalities either turn *Non-Reductive Physicalism* into a reductionist view, or one that is non-physicalist. I do not take this to show that the non-reductionist is in fact committed to zombies or the relationship to *Substance Dualism* that I have argued followed from zombies. Rather, it establishes that in so far as the non-reductionist appeals to a certain supervenience relation in the formulation of his position, his view is incoherent.

This should already have undermined the availability of appeals to supervenience by non-

reductionists. So, appealing to supervenience to distinguish *Non-Reductive Physicalism* and *Strong Property Dualism* should look bleak. However, I will make a further case against supervenience based on more traditional worries. It will be based on the requirement that the supervenience relation be explainable in a physically acceptable way. I will follow Terrence Horgan in arguing that whatever physically acceptable explanation there could be for supervenience is a reductive one. And that it is not an option for the non-reductionist to let the supervenience relation go unexplained lest supervenience be uninformative and unable to secure physicalism.

3.1 Strong Property Dualism, Zombies and Substance Dualism

The possibility of zombies is a notorious commitment of the strong property dualist. ‘Zombies’ are meant to be items physically and functionally identical to humans, but which lack conscious experience. This means that one’s so-called zombie twin will be molecule for molecule identical with oneself, as well as display the same sorts of behavior; but lack conscious experience. Since Chalmers is our paradigm strong property dualist, I will use his argument for zombies to represent the motivation for why these strange creatures are considered possible. It goes like this:

1. Zombies are conceivable.
2. If zombies are conceivable, then they are possible.
3. If zombies are possible then physicalism is false.

‘Conceivability’ here is defined in terms of ‘logical/conceptual coherence’. In other words, we can say that something is conceivable, if our description of it results in no formal contradiction of logic or contravention of the rules governing the meaning of the concepts at issue. In this way, the only restriction on whether zombies are conceivable is that the concept of a zombie presents us with no logical or conceptual contradiction. The supporter of zombies could say then that if there is no relationship of entailment between functional and physical concepts; and mental concepts, then zombies are conceivable. Conceivability so construed is then identified with possibility. For the strong property dualist, the conceivability of zombies has metaphysical significance. To avoid the apparent conceivability from remaining a fact about the relations between mental concepts and physical/functional ones, the identification of conceivability and possibility must be made. The arguments for this identification are not important in the context of this paper. What is important is that *both* the conceivability of zombies and the identification of conceivability with possibility are required for the

conclusion of the zombie argument.

One clear implication of the possibility of zombies is that physicalism is false. It seems clear that on a minimal physicalism zombies would at least not be possible. Chalmers himself takes the possibility of zombies to support the truth of *Strong Property Dualism*. With this in mind, I now present two arguments centering on the possibility of zombies revealing a problematic relationship between *Strong Property Dualism* and *Substance Dualism*.

3.1.1 *Zombies, Modality and the Bearers of Mental Properties*

The first argument aims to show that *Strong Property Dualism* is at the very least consistent with *Substance Dualism*. It begins by asking: “What bears mental properties?” This is a pertinent question because properties must have bearers.¹⁹ In this way, making clear what it is that *has* mental properties is a task that must be undertaken by all accounts of mind that make reference to mental properties. If the strong property dualist is to be a substance physicalist, it is trivial that whatever story he tells about the bearers of mental properties, that they be *physical*. The trouble for the strong property dualist then is that every physical candidate for the bearer of mental properties is the sort of thing that the zombie argument says *can* exist without mental properties. This follows logically from the premise of the zombie argument that asserts the possibility of physical identity and mental divergence.

I take this to be a problem for the strong property dualist who aspires to avoid substance dualism due to a particular view that I hold regarding the truth conditions for the sort of modal claims used in the philosophy of mind. The view could be expressed as follows:

Moderate Modal Actualism: At least some of our modal claims must be made true by something about the actual world.

This might seem like a simple claim, but it is a contested issue in philosophy whether all modal truths require truth makers to begin with, let alone that they are made true by features of the actual world. (Cameron 2008 p. 26) However, the modal claim made by the strong property dualist here seems to *require* a special relationship to the actual world. It seems clear to me that when we ask questions about the nature of the mind, we are asking questions about the nature of the actual world. When I worry about minds being physical or non-physical, I am worried about *actual* minds – my mind, your mind. So, when philosophers use modal

¹⁹ I will help myself to this sort of view, but it seems very common in philosophy to think that where there are properties, there are substances that bear them.

claims in discussions about the nature of minds, I take them to be employing tools that are somehow fit to illuminate the nature of the actual world. If these claims did not have this aim, or were unfit for this task they would be useless in matters of ontology. One straightforward way for these claims to be relevant to the actual world is to expect them to be made true by features of the actual world. Ross Cameron seems to have something like this in mind when he analyses what it is to be a possible world:

There is nothing more to a world's being a possible world, I suggest, other than that it represents the actual laws of logic, the actual mathematical truths, the actual natural kind identities, etc, as being true, and that it doesn't represent as true anything that contradicts any of these truths. In that case, to account for the fact that some world *w* is a possible world, we need only account for the fact that the actual truths concerning logic/mathematics/natural kind identity etc are represented as being true by *w*, and that nothing that is incompatible with any of these truths is represented as being true by *w*. So let *S* be the set of all and only the propositions that *w* represents as being true: whatever makes it true that *w* represents all and only the members of *S* as being true will also make it true that *w* is a possible world – and it is plausible that this is merely *w* itself. (Cameron 2008 p. 280)

The idea here is that worlds are possible in so far as they reflect and don't contradict truths about the actual world. (Cameron 2008 p. 280) So, if this is the right way to think of the relationship between modal truths and the actual world, then zombies are possible in so far as they are consistent with some set of truths about the actual world.

Cameron notes, however, that this is not to take a position on what those truths *are*. In this way, although he links modal truth to the actual world, he does so in a rather loose way – modal truths require that they reflect some truths about the actual world. This means that we would have options if we were to ask the strong property dualist *what it is* about the actual world that makes it true that zombies are possible. One option, is that the actual world is such that mental properties are not borne by any physical subjects. A zombie world, remember, is just a coherent description of the physical aspects of the actual world minus its mental aspects. Why not think that in moving between worlds what we 'leave behind' in the actual one is the mental substance that bears mental properties? This, of course, does not exclude the ontology the strong property dualist describes. Given the possibility of zombies it could very well be the case that there are only additional mental properties in the world borne by physical substance. But at the very least it makes *Substance Dualism* compatible with *Strong Property Dualism*. The strong property dualist needs to say what it is about his position that rules out *Substance Dualism*.

The trouble for the strong property dualist is that there is nothing he *could* say here short of merely stipulating that he wants to avoid substance dualism. Why *couldn't* it be the case that the reason why zombies are possible is because mental properties are not borne by any physical substance?

3.1.2 *Zombies and Ghosts*

The next argument takes the relationship between *Strong Property Dualism* and *Substance Dualism* to a new level. It aims to establish that the former entails the latter. Recall that if you are a strong property dualist, your commitment to zombies means that it is possible to have everything that is physical about the actual world but have no phenomenal experience. The same commitment entails that we could have minds in a world that is very different – physically speaking – to the actual world. On this sort of view, there is a clear sense in which the mental enjoys independence from the physical. Nevertheless it is part of the strong property dualist's narrative that whenever there are minds, there must be *some* physical aspect that grounds them. In other words, mental phenomena are not considered as able to exist free of any physical grounding whatsoever. If something like this were possible, then we might begin to doubt whether mental properties were borne by physical substances.

In a recent article, Paul Goff provides a compelling argument that if zombies are possible, then so are what he calls 'ghosts'. A ghost is the sort of thing whose nature is exhaustively mental. A ghost has no physical properties at all, just mental ones. Ghosts are different to zombies in so far as one's 'zombie twin' is physically identical to one but mentally different while one's 'ghost twin' is mentally identical while having no physical properties whatsoever. (Goff 2010 p. 119) A commitment the possibility of ghosts, then, seems to be the same as a commitment to a Cartesian version of the argument for dualism. This sort of argument clearly establishes the distinction of physical and mental substance and so cannot be a commitment of *Strong Property Dualism* if its adherents do not want to be committed to *Substance Dualism*.

The possibility of ghosts needs to be thought of along the same lines as the possibility of zombies. This means that it first requires a step claiming that we can *conceive* of ghosts and the assumption that this entails that ghosts are *possible*. Only if the standard is the same will Goff's argument work. With this in mind, what considerations could one bring against the possibility of ghosts?

One consideration would be if ghosts were inconceivable. In order to appeal to this consideration against the possibility of ghosts one would need to claim that the notion of a ghost is incoherent. To do this as a *strong property dualist*, however, one would need to claim that while it is coherent to conceive of a physical or functional duplicate of oneself that lacked phenomenal properties, it is in fact incoherent to conceive of a mental duplicate of oneself that lacked any physical or functional property whatsoever. The trouble with this sort of response, however, is that it seems to contradict the conceivability of zombies. It is precisely because there is no conceptual overlap between mental and physical or functional properties that zombies are conceivable in the first place. If there is no trace of the mental in physical or functional concepts, then there ought to be no trace of the physical or functional in *mental* concepts. To expect the failure of conceptual entailment to go only from the physical or functional to the mental but not the other way would at best be *ad hoc*. In this way, the strong property dualist cannot deny the conceivability of ghosts without denying as well the conceivability of zombies. Therefore, in so far as the strong property dualist is committed to the possibility of zombies, he is committed to the possibility of ghosts as well. (Goff 2010 p. 124-127)

Another option is to rule out ghosts not by saying that they are incoherent, but rather to appeal to *a posteriori* considerations in an effort to show that ghosts are not *possible*. He explains:

Let us say that we have decided, through reflection on zombies... that there is no a priori connection between [physical] states and conscious states. We do not yet have a reason to believe in the *metaphysical*, as opposed to merely conceptual, distinctness of physical states and conscious states.

If conscious states *turn out*, as a matter of empirical fact, to be physical states, then [the conceivability of zombies] although true, has no metaphysical significance (Goff 2010 p.126)

The idea here is that the conceivability of ghosts could be admitted, while any metaphysical implication could be blocked by appealing to *a posteriori* identifications of phenomenal states with physical states of our brains. Kripke argues that this route is available as a response to philosophers who thought that the apparent conceivability of water without H₂O meant the relation between water and H₂O is genuinely contingent. Kripke showed, however, that the apparent conceivability of water without H₂O did not reveal that there is any contingency in the relationship between water and H₂O. Kripke argued that when we think of the association between water and H₂O as contingent, we think of an epistemic counterpart of water – something that looks like water but is not H₂O. This is not to show that we can have water without H₂O, however. In this way, Kripke blocks the move from

the conceivability of water without H₂O to the idea that the relationship between water and H₂O is genuinely contingent in some metaphysical sense. In the case of water and H₂O, whenever we conceive of water without H₂O we are really only conceiving of an ‘epistemic counterpart’ of water. But if water is H₂O, then necessarily water is H₂O, so we can’t have situations in which the two come apart.

In the same way, then, one might say that whatever plausibility there is in claiming that ghosts are conceivable, they are not possible. The trouble is of course that *a posteriori* identification shows that there *is* a gap between what is conceivable and what is possible. If we are to take Kripke’s considerations about water and H₂O as a model, then there certainly is no identification of conceivability and possibility. In so far as the identification of conceivability and possibility is an essential premise in the argument for the possibility zombies, this route to objecting to ghosts is closed to the strong property dualist as well.

The strong property dualist might respond here and say that this is not a fair analogy because pain is different from water and H₂O – even according to Kripke. Kripke famously argued that *a posteriori* identifications like the identification of water and H₂O are not available in the case of certain mental states – he uses ‘pain’ as an example. Kripke argued that the contingency between pain states and states of one’s neurological system (C-fibres, say) could not be explained away as easily as it was in the case of water and H₂O. If we were to take the same strategy with pain, then we would need to say that scenarios in which it seems conceivable to have pain without C-fibres are really scenarios in which we have epistemic counterparts to pain, but really where there is no pain at all. The trouble is that we don’t seem to have anything like an epistemic counterpart to pain that is not itself pain. An epistemic counterpart to pain would be a state that appeared to be pain, but that is not in fact pain. The trouble is that appearing to be pain seems to consist minimally in *feeling* like pain. But, so Kripke’s argument goes, if something feels like pain, it *is* pain. Being a pain is just to feel a certain way. So, here one has not been able explain away the contingency between pain and C-fibres. Here the contingency is metaphysical. In this sort of case, we would not be able to identify pain with C-fibres in the way we did with water and H₂O. So whatever Kripke says about *a posteriori* identities, they do nothing to show that conceivability and possibility are not the same in the case of pain.

In the context of Goff’s argument, though, this point is unlikely to serve the interests of the strong property dualist. If it is right, then it is still the case that the conceivability of ghosts entails the possibility of ghosts. And in any case, it could be that Kripke’s claim that pains and C-fibres are contingently related is only a problem for so-called type identity theories and

not token ones. The idea here is while Kripke may have shown that type identity is out of the question, we could assume his point about the apparent unique relationship between pains and neurological states and still accept the token identity of mental states with mental tokens. The trouble is, token identity is counter to the core claims of *Strong Property Dualism*.

This, however leaves open a route to *Non-Reductive Physicalism*. Recall that I am at present trying to establish whether *Non-Reductive Physicalism* can distinguish itself from *Strong Property Dualism*. The two arguments immediately above were meant to show that *Strong Property Dualism* has a problematic relationship with *Substance Dualism* while *Non-Reductive Physicalism* did not. While the strong property dualist seems caught in a dilemma in so far as he must either admit ghosts (and thus *Substance Dualism*) or reject zombies, the non-reductionist can accept the abovementioned Kripkean considerations but remain true to his core claims. The non-reductionist here could admit that water and H₂O does not provide a model for the identification of pains and C-fibres, but go on to claim that there is still scope for a view on which mental tokens are identical to physical ones. The idea here is that Kripkean considerations are able to show that pains and C-fibres are distinct *as types*, but not that particular instances of pains are not identical to particular instances of C-fibres. Ostensibly then, the non-reductionist is not in the same boat as the strong property dualist because he still has this notion of token identity to which he can appeal. The issue of token identity is very important to the formulation of *Non-Reductive Physicalism* and ultimately to whether it is able to be a coherent physicalist position. I will leave token identity to the side for the moment and focus on the notion of supervenience as it appears in the formulation of *Non-Reductive Physicalism* and the apparent benefits that it generates in terms of distinguishing *Non-Reductive Physicalism* from *Strong Property Dualism*.

3.2 Supervenience and Non-Reductive Materialism

So far I have argued that the strong property dualist's commitment to zombies shows that the view has a problematic relationship with *Substance Dualism*. The most obvious way in which the non-reductionist can avoid these arguments is, of course, to claim that zombies are impossible. This of course is something stipulated by the non-reductionists but my concern here will be to determine if it can be substantively defended. One surefire way to do this is by appealing to a supervenience relation. The one most commonly used by non-reductionists is *Strong Supervenience*, which is defined like this:

Strong Supervenience: "Mental properties supervene on physical properties in the sense that if something instantiates any mental property *M* at time *t*, there is a

physical base property P such that the thing has P at t and necessarily anything with P at a time has M at that time.” (Kim 1998 p39)

As it stands, however, this definition of supervenience is ambiguous because it leaves a modal operator ‘necessarily’ undefined. *Strong Property Dualism* claims that the dependence between mental and physical is based on nomological necessity, which is analyzed in terms of the holding of contingent natural laws. This means that the mental depends on the physical only as a matter of contingent natural law. So, we might have worlds just like the actual world, but with different natural laws such that at those worlds there are no mental properties. This clearly can’t be the form of necessity operative in a non-reductionist characterization of his supervenience relation. There are only two popular ways of characterizing the modality of a given supervenience relation other than nomological necessity: logical and metaphysical necessity. The non-reductionist will need to choose between these two.

3.2.1 Alternate Necessities for Supervenience

Logical necessity is commonly analyzed as *conceptual* necessity – that is, a form of necessity restricted only by the rules of logic and the *meaning* of predicates and concepts. So, we might say that it is logically *impossible* to have a square circle, or a married bachelor. This is the way that *a priori* reductive physicalists describe the relation of the mental to the physical. It is clear, however, that if this is what logical necessity consists in, it cannot serve as a description of the dependence between mental and physical for the non-reductionist. It clearly entails the reduction of higher-order phenomena to lower order ones. So, the only real option for the non-reductionist is that the mental supervenes on the physical with metaphysical necessity.

The concept of metaphysical necessity, however, has received different analyses by different philosophers. My task here will be to determine whether there is a sense of metaphysical necessity that is available to the non-reductionist. One formulation comes from Chalmers. He says that what is metaphysically necessary is a function of evaluating the true application of a predicate in terms of that predicate’s secondary intension as opposed to its primary intension. The stock illustration here is ‘water’ and H_2O . The primary intension of ‘water’ is something like ‘a clear liquid substance’. The secondary intension is H_2O – what we discover this ‘clear liquid substance’ to be upon empirical investigation. On this picture it is easy to see how it is in a sense ‘possible’ to have water without H_2O – we could easily think of a situation where there is ‘a clear liquid substance’ with a chemical composition that is not H_2O . In contrast there are no worlds at which there is H_2O without water. Chalmers goes on to argue that there really is no distinction between logical and metaphysical necessity. He thinks that the

secondary intension of 'water' is just as good a candidate for the meaning of 'water' as its primary intension. What is metaphysically possible, then, is just another species of logical necessity. I won't take a position on Chalmers' analysis of metaphysical necessity here, but it is clear that if it is right then, metaphysical necessity entails a reductive relationship between mental and physical and is thus not suitable for the non-reductionist's purposes.

But let's suppose that Chalmers is wrong and the two forms of necessity are meaningfully distinct. If this were the case we could say that the primary intension of 'water' gives its meaning but its secondary intension doesn't. The secondary intension here will serve instead to show what water *is*. In this way, one might say that it may be metaphysically possible to have 'a clear liquid substance' without water but metaphysically impossible to have water without H₂O. One might wish to view 'pain' the same way. If we could, then the non-reductionist can appeal to metaphysical necessity to contrast himself with the strong property dualist. He could say that the primary intension of, 'is in pain', is one thing, its secondary intension is another. And just like we found out that H₂O is the secondary intension of 'is water', we might find that in the actual world that pain can be identified with a certain neurological structure, say firing C-fibers. And just like it is with water, he could admit a sense in which it is *possible* to have pain without C-fibers firing but still (metaphysically) impossible to have C-fibers without pain. Since zombies are the sorts of thing that *would* have C-fibers without pain, analyzing metaphysical necessity like this rules out zombies and thus troubles regarding consistency with *Substance Dualism*.

The trouble for the non-reductionist is that if he is to analyze metaphysical necessity like this, he will have to admit that the semantic properties of 'pain' are the same as 'water'. This would mean that the primary intension of 'pain' would be contingent and its secondary intension, necessary. However we seem to have good reason to think otherwise. A popular candidate for the primary intension of 'pain' is 'a state that feels such and such'. It would be strange if the way pain felt were something contingent about *pain*.²⁰ This is of course a position that some philosophers do take. Richard Brown (2010) and Christopher Hill (2012) are two important contemporary examples. They both agree that it is possible to have pain without the feeling of pain because pain just is C-fibers. I think it is safe to say, that this runs counter to *Non-Reductive Physicalism*.

In light of this, one might analyze metaphysical necessity in terms of 'essences'. This is not

²⁰ Chalmers holds that phenomenal concepts do not have distinct primary and secondary intensions, which he takes to be the reason that pain cannot turn out to be c-fibre firing in the way water can turn out to be H₂O [1996; 2002; 2009]. Taken from Goff (2010).

the same as analyzing it in terms of primary and secondary intensions, but rather, in terms of the features a thing has if it exists at all. This route has the benefit of providing a contrast between logical and metaphysical necessity in so far as a thing's essence need not be a function of the meaning of the words used to refer to it. This, I take it was the point that Kripke makes when he shows that there are *a posteriori* (non-conceptual) identities. If we took this route, we could restrict the metaphysically possible worlds to those in which things have at least their essences. On this analysis of metaphysical necessity, however, metaphysically possible worlds would include worlds where there are C-fibers without pain, since C-fibers are not part of the essence of pain.²¹ This rules out the important physicalist tenet that we cannot have physical identity without mental identity.

So, in order to avoid zombies, the non-reductionist must appeal to metaphysical necessity. But we don't yet have an analysis of this form of necessity consistent with his core claims.

3.2.2 Explaining Supervenience

Suppose, however, that contrary to what I have argued in the previous section, we have a supervenience relation in place that *is* able to distinguish the non-reductionist from the strong property dualist. Might there be considerations independent of the ones raised above that serve to undermine a non-reductionist's appeal to supervenience? In this section, I will argue that there certainly are. Whatever the non-reductionist expects from *Strong Supervenience*, he needs it to secure the impossibility of physical identity and mental difference. Trouble begins to head the way of the non-reductionists when one asks for an *explanation* of the supervenience. The requirement for such an explanation of supervenience stems from some historical considerations about the use of the word and idea of supervenience in the history of philosophy. There we see that 'supervenience' has been considered compatible with anti-physicalism. G. E. Moore countenanced a supervenience relation (although not by name) in explaining the relation between non-physical moral properties and physical ones. C.D Broad – and other British Emergentists – considered his emergent phenomena to be supervenient on physical phenomena. While this does not support a conclusion that supervenience is compatible with the falsity of physicalism, it does generate an obligation on the part of physicalists who wish to appeal to supervenience to define in such a way that rules out views about the mental that are anti-physicalist

Terrence Horgan's paper is an important starting point for any discussion about the

²¹ This objection to the availability of metaphysical supervenience is based on the assumption, of course, that the feeling of pain *exhausts* the nature of pain.

explainability of supervenience. He too worries about an unexplained supervenience relation. He argues that if supervenience is to be incorporated into any physicalist account of mind, it needs to be “robustly explainable in a physicalistically acceptable way”. I’ll call this *Horgan’s Constraint* and define it like this:

Horgan’s Constraint: Any genuinely physicalist metaphysics should countenance ontological inter-level supervenience relations only if they are robustly explainable in a physicalistically acceptable way. (Wilson 2012)

When going about giving a robust explanation, Horgan thinks that three questions need to be answered:

The Standpoint Question: What sorts of facts, over and above physical facts and physical laws, could combine with physical facts and laws to yield materialistically kosher explanations of inter-level supervenience relations, and why would it be kosher to cite such facts in these explanations?

The Target Question: What facts specifically need explaining in order to explain a given inter-level supervenience relation, and why would a materialistic explanation of *these* facts constitute an explanation of that supervenience relation?

The Resource Question: Do there exist adequate explanatory resources to provide such explanations? (Horgan 1993 p. 578)

The first two requirements are related. Before we determine whether there is a robust explanation for mental phenomena, we must first have in mind what those phenomena are. The answer to this question is the answer to the target question. The standpoint question involves deciding what sorts of fact would combine with the basic facts of physics to provide a robust explanation for supervenience. The idea here is that facts about fundamental physics could combine with facts about phenomena that are uncontroversially physical – like chemical, biological and neurological facts. The resource question then involves our ability, given the answers we have to the previous questions, to provide an explanation of the target facts in terms of fundamental physical facts in concert with other physical facts. He thinks that our current explanation of the phenomenon of ‘liquidity’ is a paradigm case of robust explanation. In the case of liquidity, we seem to be able to give an explanation of the facts definitive of liquidity in term of other perfectly physical facts. (Horgan 1993 p. 578-580)

Horgan refers to attempts at achieving what seems achievable for ‘liquidity’ for mental

phenomena. He calls these, ‘naturalizing projects’. He says that each begins rightly by attempting an answer for the target question. That is, they first aim to get clear on the phenomena we’re trying to give a physically acceptable explanation *for*. Importantly, the task of defining the target question involves defining the ‘definitive’ features of the mind so that they *can* be susceptible to a physical explanation. (Horgan 1993 p. 579) He mentions functionalism as one answer. Functionalism has the benefit of defining targets that we seem to have the resources to explain robustly. The trouble is, that in giving the definitive features of mental phenomena in functional terms, one has in effect produced a reductionist view of mental phenomena.²² There seems to be a dilemma here. In so far as the target facts about the mental are considered definitive of the mental, if they are the sort that are *able* to combine with physically acceptable claims we have a reductive conception of the mental. The only sorts of target facts that we seem to have the resources to explain are the sorts of target facts that we’d get only if we had a reductive view of the mind! He writes about naturalizing projects:

Their goal is to give a tractable specification, in non-intentional and non-mental vocabulary (although not necessarily in the vocabulary of physics), of sufficient conditions (or sufficient and necessary conditions) for the instantiation of mental properties. To the extent that this could be done, it would pave the way for physicalistic explanations of supervenience connections.

In so far as an explanation of the supervenience relation must involve recourse to this sort of process, supervenience becomes a reductive relation. This is problematic for two reasons. The first is that one of the virtues of supervenience in the first place was that it provided a non-reductive, yet physically acceptable account of the relation between mental and physical. The second is that the non-reductionist clearly can no longer appeal to supervenience as part of the formulation of his position.

In a recent paper, Jessica Wilson makes a point that seems to make the robust explanation of supervenience even more difficult to incorporate in a non-reductionist framework. The point is that *Horgan’s Constraint* must rule out the possibility of higher-level phenomena having causal powers that are different from physical phenomena. She points out that this is a minimal condition for physicalism. What she recommends is that we specify in our conditions for robust explanation that it rule out a difference in causal powers between higher and lower levels. This seems rather sensible if not perhaps obvious. But if our explanation of the supervenience relation shows that the causal powers of higher-level phenomena just are the causal powers of the lower level one, we seem to have a reductive view. What sense – except

²² Think back to chapter one and functional reduction.

maybe conceptual – is it still appropriate to say that there is a distinction between mental and physical if their causal powers are the same? I am with Kim in thinking that: “...unless mentality made causal contributions that are genuinely novel, the claim that it is a distinct and irreducible phenomenon over and beyond physical-biological phenomena would be hollow and empty.” (Kim 1993 p. 204)

This dilemma is almost a replica of the one at the center of my argument that *Non-Reductive Physicalism* has a problem with mental causation. There I argued that either it is the case that non-reductionists admit that mental properties – in so far as they are distinct from physical ones – possess causal powers that are distinct from physical ones or that they have identical causal powers. In the former case, we have the falsity of physicalism. In the latter case we have a position on which it hardly seems that mental and physical properties are distinct.

3.2.3 Therefore, No Supervenience

Therefore, the non-reductionist cannot appeal to supervenience. Besides logical, nomological and metaphysical necessity, there are no other notions of necessity that could provide an analysis of modal strength of his supervenience relation – without conflating his view with either the strong property dualist or the reductionist, that is. Moreover, the task of explaining supervenience has a similar result. Supervenience explained is reductive and supervenience unexplained is uninformative and possibly consistent with the falsity of physicalism. This is a problem for the non-reductionist because it means that he can't distinguish himself from *Strong Property Dualism* by appeal to supervenience. This does not necessarily show that *Non-Reductive Physicalism* is committed to *Strong Property Dualism*, or that *Strong Property Dualism's* problems are *Non-Reductive Physicalism's* problems but it does show that in so far as *Non-Reductive Physicalism* relies essentially on an appeal to a supervenience relation in the formulation of its position and in order to distinguish itself from others, *Non-Reductive Physicalism* has not been shown to be a distinctive and coherent position.

3.3 Token Identity

Other than supervenience, another popular way that non-reductionists have tried to set themselves apart is by an appeal to token identity.²³ In this paper, I will define token identity like this:

²³ I will assume that *Token Identity* is incompatible with *Strong Property Dualism* and *Substance Dualism*. If I am wrong about this, *Token Identity* is excluded from distinguishing *Non-Reductive Physicalism* from *Strong Property Dualism* right from the start.

Token Identity: Although mental and physical types are not, particular instances of mental states (mental tokens) are necessarily identical to particular physical states (physical tokens). (Schneider 2011 p. 1)

I do not think this option is available to the non-reductionist. Susan Schneider presents a simple, but compelling case in this regard. She argues that *Token Identity* is inconsistent with the rejection of type identity. She makes her case with the help of some claims about the nature of universals. She points out that philosophers have had different ideas about the nature of universals. She quotes Armstrong, who gives a helpful survey of the landscape:

Properties... are thought of by some philosophers as having a nature that is self-contained, distinct from the powers that they bestow. We shall call this position Categoricalism. Others think of them as having a nature that essentially looks beyond the particulars they qualify, outward to potential interactions with further particulars, and where this nature is exhausted by these potential interactions. This view may be called Dispositionalism. As one might naturally expect, there is a Two-sided View which holds that properties have, essentially, both a categorical and power side to their nature.⁶

She thinks that whatever your view in this regard, if you think that properties are universals, you can't appeal to token identity. Her argument stems from the rather simple claim that a token's membership to a particular type ought to be explainable by virtue of some feature of the token that is unique to its type. The options are given by Armstrong immediately above: either tokens belong to their types due to a unique categorical, or dispositional nature or a nature that is a combination of categorical and dispositional. The trouble, however, is that the rejection of type identity means that tokens must differ in terms of the features by virtue of which they belong to their types. Pains, for instance belong to some type in virtue of some feature unique to pains and C-fibers belong to some physical type in virtue of some unique features of *theirs*. This means that mental and physical tokens by definition have at least one disparate feature – the feature that accounts for their membership to distinct types. This point could be made whatever position you took on the nature of universals. The feature that explains type membership could be a categorical property, a dispositional one or some combination.

Here's the argument more formally:

1. Tokens have the same nature as their types.

2. So, mental tokens have the nature of their type and physical tokens have the categorical nature of *their* types.
3. Type identity is rejected by *Non-Reductive Physicalism*.
4. So, mental and physical tokens have distinct natures because their types have different natures.
5. By 2, 3 and 4, mental and physical tokens can't be token identical because they have distinct natures.²⁴ (Schneider 2011 p. 3-5)

Schneider considers some responses. She says that we can't say that tokens have different natures to their types because then there would be no sense in saying that they belong to that type. Nor can we say that mental and physical tokens have the same nature. This would conflict with its rejection of type identity. Appeals to *Token Identity* by non-reductionists are therefore ruled out because *Token Identity* is incoherent. (Schneider 2011 p. 3-5)

There might be a worry that one would be endorsing type identity by endorsing this particular critique of token identity. I don't think this is an argument for type identity, though. It simply has as one of its premises that type identity is false. This is because Schneider is taking the fight to the non-reductionist, not because she is making a case *for* the type physicalist. The thrust of her argument seems to be that what the non-reductionist wants to achieve by appealing to token identity does not cohere with the denial of type identity. This is not the same as arguing for type identity.

3.4 A New Spectrum

So far, I think I have undermined *Non-Reductive Physicalism* as a viable account of mind. Whether it is *Non-Reductive Physicalism's* troubles with mental causation or difficulties articulating a notion of supervenience that coheres with its core claims, I think I have provided a strong case against it. My suggestion at this point in this paper is that we drop it from the spectrum of traditional views set out in my introduction and in chapter one. Moreover, I hope to have shown that at the very least the lines between *Strong Property Dualism* and *Substance Dualism* are blurred, if not that the latter entails the former. If this is the case then along with *Non-Reductive Physicalism*, we ought to drop *Strong Property*

²⁴ Note that this argument works against views on which property natures are combinations of categorical and dispositional properties as well. If it is problematic to appeal to either categorical or dispositional property natures, then clearly views that combine these two unavailable positions will be unavailable as well.

Dualism from the spectrum also on the charge of incoherence.

This leaves a rather bleak landscape for philosophers of mind to choose from. In so far as one values physicalism, one seems forced to persevere with a form of reduction. In so far as one values *non-reduction*, one seems to be required to adopt a rather extreme version of non-physicalism. Indeed, I have done nothing to show that either sort of position is ruled out. Readers who do not share my bias towards physicalism and/or against reduction are free to take the findings of the previous chapters to form a prima facie case for the plausibility of either reduction or non-physicalism. The rejection of this bias notwithstanding, a factor clearly relevant to one's reaction to the previous two chapters will depend on whether one thinks that there is available a physicalist position that is neither like reduction nor like *Non-Reductive Physicalism*. In the next chapter I will set out an alternative that reopens hopes for a physicalist account of mental phenomena.

Chapter Four: An Alternative

In this chapter, I want to explore the possibility of an account of mind not mentioned on the original spectrum set out in chapter one. The view I will discuss in this chapter rejects (P) and the metaphysical framework that comes with it. In this way, it is free to explore ways of being physicalist about the mind without being committed to reduction. The view is based on one that John Heil describes (2003 and 2012). An outright endorsement of this sort of view will be beyond the scope of this paper. All I will do is set out the central tenets of the view and show how – if it *can* be defended – it avoids the problem of mental causation and the problem of zombies. While not an outright defense of Heil’s view, the ability to avoid these problems is a significant virtue of his view that makes his account worthy of further consideration.

Heil’s view does, however, take on several significant challenges. I will focus on three. Each has to do with the apparent inability of his account of mind to explain certain data. These include: That distinct items – sometimes physically diverse – can agree in mental attribute; that ‘realism’ is true about mental states;²⁵ and that some of our mental states have phenomenal features. Heil clearly *intends* for his view to be realist about mental states and to be able to make sense of the fact that physically diverse items can agree in mental attribute. However, I will argue that whether he has the resources to make good on these intentions depends on his ability to account for the phenomenal properties of experience within an ontology of *only* physical properties. Heil’s most formidable challenge in this regard will be presented as accounting for the apparent ‘subjective’ nature of phenomenal experience, contrasted with the ‘objectivity’ of the physical properties.

Heil offers two ways in which we can accommodate the subjective aspects of experience in a world where there are only physical properties. The first strategy combines the identification of so-called ‘categorical’ and ‘dispositional’ properties with the identification of the subjective aspects of experience with categorical properties. The second is a modified version of representationalism about subjective properties. Heil claims that we can get a long way away from having to admit that qualia pose intractable problems for a view like his by appealing to classic representational accounts of mind on which qualia are apparently exhausted by the properties of objects that we represent in experience as opposed to properties of the experience itself. Heil stops short of endorsing this sort of account of phenomenal experience because he thinks that there is, contrary to classic representationalist

²⁵ By realism here I mean that talk about mental states is talk about objective features of the world. So-called anti-realists or ‘eliminativists’ about mental states say, on the contrary that talk of mental states fails to refer to any feature of the mind independent world.

views, indeed non-representational content of that forms a part of the content of phenomenal experience. He endeavors to account for this non-representational content, however, by appealing to the representational *medium* of experiences, which he thinks in turn can be described in entirely physicalistic terms.

All of the above strategies are *deflationary* responses to the problem that the apparent subjectivity of phenomenal states poses for an ontology on which there are only physical properties. In each case, Heil is in effect saying that there is no real problem of subjectivity for physicalist accounts of mind because subjectivity can be explained away.

I will not claim that Heil has provided knockdown arguments for either account of phenomenal properties that he puts forward. Nor will I claim that they can be decisively refuted. I will argue, however, that if one is to reject Heil's deflationary strategies – indeed any deflationary strategy – a *non-physicalist* position is the most appropriate alternative. I will argue that denying that deflationary approaches to subjectivity means that one objects *in principle* to the possibility that subjectivity can be explained by appeal to physical, objective, properties. The seemingly intractable nature of the distinction between subjective and objective properties seems to force any good physicalist into some kind of strategy where less is made of the apparent subjectivity of experience. In this way, I hope to place in the way of *Non-Reductive Physicalism* yet another dilemma. In so far as the non-reductionist admits that the divide between subjective and objective properties presents an intractable problem for physicalism, he is not really a physicalist after all. And in so far as he admits that there is in principle a way to explain the subjective aspects of experience in terms of objective physical properties, he seems to be endorsing a position where it doesn't make all that much sense to say that there is an ontologically interesting distinction between mental and physical properties.

4.1 Heil's View

4.1.1 The Replaceability of (P)

Heil rejects (P) in so far as it includes (P₂) and (P₃). Heil still wants to preserve the referential role of mental predicates. We could say then that he accepts at least (P₁) but not (P) in so far as it is a combination of (P₁), (P₂) and (P₃). If Heil can reject (P₂) and (P₃), it would be a significant benefit for his view because it lifts (R).²⁶ If we do not think that predicate identity

²⁶ Recall:

is property identity and that identical predicates refer to identical properties, then we are not limited to choosing between reduction, higher-level properties or eliminativism. The path would be open to a view on which mental properties just are physical ones without having to be accompanied by reduction. Whether (P₂) and (P₃) are in fact replaceable, however, will depend on whether we *can* get on without them. These two principles have long been associated with significant theoretical benefit. Many philosophers have thought they play a central role in the explanation of attribute agreement and the avoidance of eliminativism. Heil does not think that he loses any theoretical benefit on his view, however. Whether he is able to defend this will be the subject of later sections. I will first set out the central tenets of his view.

4.1.2 Substance, Property and Truth Making

Heil views the ontological categories of ‘substance’ and ‘property’ as complimentary. Substances are the bearers of properties. He thinks that it cannot be the case that we can have substances without properties nor properties without substances. (Heil 2012 p. 4) This much about Heil’s conception of substance and property is uncontroversial. Some eyebrows might be raised, however, when Heil comes to talking about the ontological/metaphysical nature of substances and properties. For Heil, substances are necessarily simple if they are to be the sorts of things that bear properties. (Heil 2012 p. 5) Simplicity here is defined in terms of lacking *substantial* parts – as opposed to spatial and temporal parts. This is quite a counter-intuitive claim because it entails that most observable objects are not really substances at all. Tables, chairs, tomatoes all are not substances, strictly speaking, because they are mereologically complex. In so far as they are substances at all for Heil, they are substances ‘by courtesy’, substances ‘loosely speaking’. (Heil 2012 p. 5) Heil does not take a position on what the fundamental simples substances *are*. He leaves answering that question up to physics. Whatever physics tells us, however, Heil is quite clear that he does not intend this conception of substance to be *eliminativist* about complex objects. Truths about complex objects, he claims are truths about arrangements of simple substance and therefore still have to do with objective features of the universe.

What of properties on this view? Properties for Heil are *modes* of substances, not universals.

(R) Where *Gs* are presumed to be uncontroversial items—those posited by the physical sciences, for instance—and *Fs* are putatively higher-level items; if talk of *Fs* cannot be analysed, paraphrased, wholly decomposed into talk of *Gs*, either *Fs* are distinct from *Gs* or there are no *Fs*.¹ (Heil 2003 p 51)

Since simple substances are the only fit bearer for properties, the only *genuine properties* for Heil are the properties of simple substances. Others are the modes of arrangements of simple substances. And just like it was with complex substances themselves, modes of complex substances are not genuine properties, but properties by courtesy, properties ‘loosely speaking’. Again, doing away with properties that are not properties of simples, Heil does not want to endorse eliminativism about the properties of complex objects. For Heil, predicate terms still have as their referents ‘properties’ and still successfully refer to them as perfectly objective features of the universe. The only difference between this and the traditional views is that the ontology of the referent is defined differently they are modes of substances, not universals. He writes:

[T]o say that a billiard ball is not a genuine substance or that the billiard ball’s sphericity or redness are not genuine properties is not to say that there are no billiard balls or that it is false that this billiard ball is red and spherical. Truthmakers—truthmakers— for judgements about billiard balls and their colours or shapes are particular arrangements of the fundamental substances. (Heil 2012)

E.J Lowe usefully describes Heil’s view as ‘truthmaker realism’. He describes it like this:

Heil endorses a view that might aptly be called *truthmaker realism*. According to it, what determines whether one is a metaphysical realist with regard to some subject matter is one’s opinion as to whether statements about that subject matter are objectively true or false, in the sense of having a mind-independent truthmaker – where a truthmaker is construed as being mind-independent if its existence and nature don’t depend upon its being *thought about* by anyone. (p. 59).

In this way, we see that Heil still expects there to be a relationship between our language and the world, even though it might not be the sort envisaged by (P₂) and (P₃). With those principles presupposed, one could determine from the truth itself what counted as the truth makers for that truth. This is because those principles entail that an identity of predicates is the same as an identity of that to which the predicates refer.

On this picture, there really are only physical properties. Mental properties, strictly speaking, don’t exist. Their fate is the same as all apparent properties of mereologically complex substances – they are properties by courtesy, properties only loosely speaking. (Heil 2012) Heil claims that any difference between mental and physical is a difference in concept, not in the world. In this way, he is not like the non-reductionist, nor like the strong property dualist. In so far as he admits the distinction between mental and physical, he will formulate it not in terms of *ontological* categories but rather categories of description.

However, this is not to say that Heil accepts a form of reduction. In so far as he admits to a difference in the meaning of mental and physical predicates, he does not align himself with *a priori* physicalists who claim that conceptual analysis is possible for mental terms. He rejects theoretical *a posteriori* reduction because he admits that there is not a single predicate that is co-extensive with any mental one. This is a fairly standard rejection of theoretical reduction based on multiple realizability. Forms of theoretical reduction that reduce the mental to the physical by means of identifying mental predicates with (possibility infinite) disjunctive predicates are rejected as hopeless. (Heil 2012)

4.1.3 Categoricalism and Dispositionalism about Properties

Another important feature of Heil's position is the view he takes on property natures. Earlier, when discussing non-reductionist appeals to token identity, I mentioned the idea that properties had *natures*. There I quoted Armstrong who helped survey the landscape of philosophical positions in this regard. I think the passage deserves a second look here. Armstrong writes:

Properties... are thought of by some philosophers as having a nature that is self-contained, distinct from the powers that they bestow. We shall call this position Categoricalism. Others think of them as having a nature that essentially looks beyond the particulars they qualify, outward to potential interactions with further particulars, and where this nature is exhausted by these potential interactions. This view may be called Dispositionalism. As one might naturally expect, there is a Two-sided View which holds that properties have, essentially, both a categorical and power side to their nature.⁶ (Armstrong)

Some philosophers take the first option and view property natures to be exhaustively dispositional. (Shoemaker 1980) On this sort of view, there is nothing more to the nature of a property than its dispositional profile – the way it interacts with other properties and the causal powers that are associated with it. A more common view, however is to consider that “every disposition must have a categorical basis in virtue of which” it has the dispositions it in fact has. (Mumford 1998) This is a very large area of philosophy that I cannot hope to treat with any adequate detail here. I will merely assume here that at least on the face of it, a view on which properties have dispositional and categorical natures is the most plausible. From this starting point, we can ask about the relationship between categorical and dispositional properties. Some philosophers think that it is contingent in some metaphysically significant

sense. On this sort of view, there are possible worlds that are identical to the actual world in terms of their categorical properties and are dispositionally different, or vice versa.

Heil's view is different. He thinks that where we invoke categorical property natures, these cannot be distinguished in any metaphysical sense from dispositional properties. Heil thinks that dispositional and categorical features of property natures are *identical*. From this it follows that the relation between categorical and dispositional properties is *necessary*. I'll refer to this claim as Heil's *Primary Identity Thesis* and define it like this:

Primary Identity Thesis: categorical and dispositional features of properties are identical.

Crucial to Heil's overall philosophical program is a modification he makes to philosophical discussions about categorical and dispositional properties. He notes that it has become so common in philosophy to assume that there is a distinction between categorical and dispositional that the term 'categorical' has become synonymous with 'non-dispositional'. To remedy any confusion, he renames 'categorical properties' as 'qualitative properties' or 'qualities'. Although Heil tries to sell it that way, the move here is not merely a terminological development. Heil goes on to identify the phenomenal properties of experience with these so-called 'qualities'. That is, for Heil, qualia *are* categorical properties. We might call this Heil's *Secondary Identity Thesis* and define it like this:

Secondary Identity Thesis: The properties of phenomenal experience are categorical properties.

As we will see, both these identity theses work together to ensure that an immediate consequence of accepting Heil's ontology is that zombies are impossible. This would be a significant virtue of his view given how problematic being unable to avoid zombies is for non-reductionists. The combination of these two theses also helps Heil in constructing a deflationary account of the subjectivity of phenomenal experience. Whether these two identity theses are plausible will be an issue I pick up in a later section. For now, I will simply assume them in order to determine what benefits may flow from their plausible defense.

4.2 Immediate Benefits of Heil's View

4.2.1 Mental Causation and Zombies

One immediate benefit of Heil's view is that it stops the problem of mental causation before it even arises. It's not quite a *resolution* of the problem as much as its *dissolution*. Heil thinks that once we have got the ontology of mental and physical right – namely that physical predicates do not refer to higher-level mental properties – the problem of mental causation as it arose for non-reductionists simply falls away.

Recall my example of the problem of mental causation for the non-reductionist, which involved the causal efficacy of higher-level mental properties *m* in the instantiation of the physical property *p*. I'll mention again the property instantiations for the sake of clarity:

M: 'my desire for fresh air'

P: 'my rising and going for a walk.'

On Heil's view, it is suggested that we view *M* here not as a higher-level property, but rather as a *predicate* that has as its referent a particular mode of an arrangement of simple physical substances. If we took '*M*' to function in this way, then we can offer a substantial review of the ontology of the causal sequence involving my desire for fresh air causing my rising and going for a walk. On Heil's picture, the truthmaker of '*M*' would be '*P**', the apparent physical state sufficient for bringing about *P* given *Closure*. If this were the case, then truthmakers for the sentence '*M* causes *P*' would just be that '*P* causes *P**'. If this were the case, the only causal sequence in the vicinity is one involving the truthmaker of a mental predicate and a physical event, both of which are physical. There simply is no question about the causal efficacy of the referents of mental predicates on this view.

There may be worries here that Heil has collapsed the mental into the physical in a problematic way. If all the causal work here is being done by lower level physical phenomena, then in what sense are our mental states relevant anymore? The important point to notice here is that Heil's recommendation is that we do away with higher-level *properties*, not necessarily higher level *predicates*. To the extent that Heil rejects conceptual equivalence between mental and physical, he accepts that there is still relevance for mental concepts and predicates in the explanation of our behavior. In this way there is at least a route open to Heil to secure relevance for mental predicates and concepts in our explanation of events involving mental and physical events. Thinking that we must inevitably lose mental predicates and concepts once we lose higher-level mental properties is a result of presupposing (P).²⁷

²⁷ Heil does, however take on the additional challenge here of showing how it is what we can have a conceptual difference between mental and physical predicates without a difference in their referents. I will not attempt an answer to this challenge, however.

Another apparent benefit that is supposed to flow from Heil's view is that zombies are not possible. It follows from Heil's *Primary Identity Thesis* that there is a *necessary* relation between categorical and dispositional properties. And it follows from his *Secondary Identity Thesis* that qualia *are* categorical properties. Together these two theses block the possibility of zombie worlds because zombie worlds are subsequently worlds in which we have dispositional identity and categorical variance. In this way, if one is to allow that zombie worlds are possible, one needs to first assume that categorical and dispositional features of properties are contingently related. There is clear evidence in Heil's writings that this is where he thinks the avoidance of zombies ought to lie.

He writes:

A Zombie is a being precisely similar to a conscious being with respect to its dispositionalities but lacking consciousness—where consciousness is taken to be a qualitative accompaniment of certain properties, properties in virtue of which a being has a compliment of causal powers or dispositionalities. But if we accept the identity thesis, then sameness of causal powers—exact dispositional similarity—yields qualitative sameness. Zombies, on this view are impossible. (Heil online entry)

Heil will need, of course, to admit – in line with his rejection of the conceptual equivalence between mental and physical vocabulary – that zombies are at least conceivable. But in so far as he accepts both identity theses, his challenge to zombies enters at the level of their apparent *possibility*. The possibility of zombies is ruled out by taking what Heil thinks is the right view on property natures.

Whether these are in fact benefits of Heil's view will depend on whether Heil's ontology can be substantively defended, however. In his various writings, he claims that his ontology is defensible on grounds independent of issues in the philosophy of mind. So, whether Heil's view is worth taking seriously might also be determined on the bases of these independent considerations. However, I have left out his arguments for his ontological theses because my main concern in this paper is to see what impact they have in the philosophy of mind if they are true. For the rest of this chapter, I will deal with the apparent difficulties of Heil's austere ontology faces in so far as he takes it to be applicable in the philosophy of mind particularly.

4.3 The Avoidance of Eliminativism

The first apparent problem is that contrary to his best intentions, Heil is unable to secure realism for mental states. That is, Heil is really committed to *eliminativism* about the mind once we consider his view carefully. To be an eliminativist about x, broadly, is to claim that our talk about a certain phenomenon fails to refer to anything in the world. An uncontroversial example of eliminativism in this sense would be eliminativism about Santa Claus. To be an eliminativist about Santa Claus is to say that there is nothing in the world that would act as the referent of the name, 'Santa Claus'. Most of us end up holding this sort of position because at some point in our lives we have learned that there is nothing at all like who Santa Claus is supposed to be in the actual world for the name 'Santa Claus' to refer to. Similarly, to be an eliminativist about *the mental* would be to say that the world lacks referents for mental predicates.

We have seen, however, that this is not the way in which Heil formulates his view. Heil still clearly intends for mental predicates to successfully refer to aspects of mind independent reality. On his view, they refer to *modes of substances*. Does this secure realism about mental states though? It certainly shows that Heil's view is ostensibly not opposed to the idea that mental terms successfully refer. The only way to challenge Heil's view from this starting point is to appeal to worries about whether modes of physical substances *could* be all we need to account for the truth of mental predications – particularly ones about phenomenal experience. Philosophers and philosophically minded psychologists have long expressed worries that views on which there are only physical properties 'leave out' some important aspect of phenomenal experience. Heil quotes an important 20th century physiologist who worries about this in great detail.

I see the sun; the eyes trained in a certain direction entrap a tiny packet of solar radiation covering certain wave-lengths emitted from the sun rather less than 10 minutes earlier. This radiation is condensed to a circular patch on the retina and generates a photo-chemical reaction, which in turn excites nerve-threads which relay their excitation to certain parts of the brain, eventually to areas in the brain-cortex. From the retina to the brain the medium of propagation is wholly nervous; that is, the reaction can be subsumed as electrical. Some of this electrical reaction generated in the eye does not reach the brain-cortex but diverges by a side-path to nerve-threads which relay it to a small muscle, which by contracting prevents excess of light attaining the retina. The chain of events stretching from the sun's radiation entering the eye to, on the one hand, the contraction of the pupillary muscle, and on the other to the electrical disturbances in the brain-cortex are all straightforward steps in a sequence of physical 'causation', such as, thanks to science, are intelligible. But in the second serial chain there follows on or attends, the stage of brain-cortex reaction an event or set of events quite inexplicable to us, which both as to themselves and as to the causal tie between them and

what preceded them science does not help us; a set of events seemingly incommensurable with any of the events leading up to it. The self ‘sees’ the sun; it senses a two-dimensional disk of brightness, and overhead shaped as a rather flattened dome, coping the self and a hundred other visual things as well. Of hint that this scene is in the head there is none. . (Sherrington 1906, xx–xxi) (Heil 2012 p. 224-5)

The idea here is that if we had to endorse Heil’s austere ontology, then something about the nature of phenomenal experience would necessarily be ‘left out’. If this worry is warranted, then if we assume Heil’s ontology, then much of what Sherrington says about this experience is, strictly speaking, *false*. Just like in the case of Santa Claus – if we become convinced that Heil is right about there being only physical properties – we have learnt that there is nothing in the world that answers to our talk of phenomenal consciousness. This of course is an unsatisfactory result for Heil. He needs to come up with a characterization of phenomenal states *as physical states* that is able to calm worries that he has left out something essential about phenomenal experience. It is not enough for him merely to assert that on his view mental predicates still successfully refer.

4.4 Attribute Agreement

Another apparent difficulty of Heil’s view is its apparent inability to explain attribute agreement. Attribute agreement is a phenomenon that occurs all the time. It happens whenever we ascribe the same property to two distinct items – for example, the sphericity of two marbles, or the blueness of two cars. However, there is an interesting form of it in the philosophy of mind. It is a widely accepted datum that items with radically diverse physical properties can agree in mental attribute. This is the so-called multiple realizability, defined in chapter one. Universalism about properties seems to have an advantage over Heil’s view in so far as we think that *Multiple Realizability* is intuitively true. If we are universalists about mental properties, then we have a ready explanation for the intuitive truth of multiple-realizability – each item agreeing in mental attribute has a numerically identical universal wholly present in it that acts as the referent of the mental predicate. Here, attribute agreement is explained by “the strict identity of universals”. (Heil 2012) Heil seems to be at a disadvantage here because he denies the existence of universals and thus loses any appeal to them in the explanation of attribute agreement. On his view, there could well be *nothing* in common between two items that agree in mental attribute.

Heil seems to think that the very claim that there is a *problem* for his view regarding attribute agreement is because (P) is already presupposed as giving the standard for an explanation of

attribute agreement. To demand that (P) is the only way to make sense of attribute agreement seems to require that we take the nature of the truth maker to match the nature of the truth in a very blunt way – the requirement for a general truth maker is drawn from the apparent generality of the truth. He writes:

Although there is a place for talk of universals, truthmakers for such talk are fully particular. Properties are *modes*, particular ways particular substances are. One billiard ball's sphericity is distinct from another's... We have general truths—about sphericity and redness, for instance—but truthmakers for these truths are particular ways the universe is, particular ways particular substances are. Generality is an important and irreplaceable feature of representations. Theories as standardly formulated consist exclusively of general assertions. But truthmakers for general truths—general truths that have truthmakers—are particular ways the universe is. The universe includes no general or universal *entities*. (Heil 2012 p. 8-9)

The idea here is that while general *terms* are unavoidable, there is no reason to expect that there be features in the world that are themselves general – general *entities* – to act as the truthmakers for general truths. Unless of course that is what you have decided before the discussion begins. What might be occurring here is a clash of intuitions. One side is assuming (P) at the start, the other is denying it. In this context we ought to expect from Heil a positive account of attribute agreement so that there is no impression that he is avoiding the issue.

Heil's positive position on the problem of attribute agreement is to appeal to a relationship between items that agree in attribute that is weaker than strict identity. He calls it 'exact similarity'. (Heil 2012) This sort of view has historically been criticized by defenders of universals for the fact that the similarity relation is brute and unexplainable. The accusation is not unwarranted given that most philosophers who admit to an ontology that rejects universals – including Heil – admit that their similarity relation is brute. In the absence of universals, it is claimed that what it is for two objects to be spherical, say, is just that they are *spherical*. This is where the explanation for attribute agreement bottoms out if we reject universals. (Goff 2007 p. 53) This in turn makes attribute agreement itself a brute phenomenon. Universalists on the other hand seem to have a view on which attribute agreement is not brute. Agreements of attribute are explained by appealing to the strict identity of universals. In so far as attribute agreement seems to be a phenomenon for which we should be able to *provide* an explanation, Heil finds himself in some trouble.

I have doubts about how deep this objection goes against Heil, though – or any view that rejects universals for that matter. Remember, Heil doesn't reject *properties* outright. He still

thinks that there *are* properties, but that they are modes of simple substances. Heil makes it quite clear that his is a two-category ontology – one that involves substance *and* property. The only difference is that his properties are modes of simple substances as opposed to universals. Paul Goff, writing in defense of ontologies that reject universals makes the point that although there is no explanation of similarity *in terms of universals* available on a view like Heil's, there still is an important sense in which attribute agreement can be explained in their absence. Take for example two items that are spherical – two marbles, for instance. Goff contends that it is still open to philosophers who reject universals to explain their similarity by pointing out that each marble has “all spatial points on their surfaces equidistant from their centers.” (Goff 2007 p. 54) This seems like a perfectly informative and acceptable explanation of why it is that the two marbles are both spherical. It would certainly be odd to say that explanations such as these are *limited* only to those who adhere to an ontology of universals. Where properties are construed as modes, one seems permitted to refer to more fundamental modes of objects to explain higher-level modes.

In this way, it is open to Heil to claim that he can explain similarity by appealing to *properties*, just not properties construed as universals. Ultimately for Heil the shared sphericity – as a mode of arrangements of simple substances – of distinct items can be cashed out in terms of *other* modes of simple substances. In this way, I think that Heil can mitigate a great deal of the skepticism surrounding his appeals to similarity instead of strict identity in the explanation of attribute agreement. Trouble arises, however, in cases where the agreement in attribute is mental. While we might be able to give informative explanations about phenomena like sphericity, it seems unlikely that we will be able to give the same sort of explanation for mental phenomena. We can see *how* it is that certain properties of marbles make it so that they are spherical (and not cuboid, for instance). But we seem to lack similar explanations for phenomenal states. Sherrington raises this worry soon after he worries that an ontology of only physical properties leaves out something essential about phenomenal experience. He writes:

Vision is saturated with this strange property called ‘projection’, the unargued inference that what it sees is at a ‘distance’ from the seeing ‘self’. Enough has been said to stress that in the sequence of events a step is reached where a physical situation in the brain leads to a psychical, which however contains no hint of the brain or any other bodily part. (Sherrington 1906, xx–xxi) (Heil 2012 p. 224-5)

This is an early statement of what has become known as the problem of the ‘explanatory gap’ between mental and physical properties. This problem is related to but distinct from the worry

that countenancing only physical properties is to leave out something essential to phenomenal experience. The apparent problem of explanatory gaps between the mental and physical is that *given* that we admit there is a phenomenal aspect to experience, how might we account for it in terms of physical processes? The problem, like the problem of physical properties leaving something essential about phenomenal experience out of our description of the world, once again is due to the apparent subjectivity of phenomenal experience and the objectivity of physical properties. How is it possible to explain essentially subjective properties in terms of objective ones?

4.5 *Heil and Qualia*

So, it seems that Heil's ability to avoid eliminativism and explain attribute agreement comes down to whether he can account for qualia on his austere ontology. Not having an account of phenomenal properties *as physical properties* has stymied Heil's efforts to prove himself as a realist about mental states. His proposed solution to attribute agreement seemed to founder on the apparent explanatory gap between phenomenal properties and physical ones. Heil proposes two strategies for accommodating subjectivity in a world of objective physical properties. If these strategies are partial to a plausible defense, then Heil is able to show that nothing essential to phenomenal experience is 'left out' given his ontology. He will also dissolve the explanatory gap by providing *and* that the problem of the explanatory gap will be dissolved as well.

4.5.1 *Categorical and Dispositional Properties*

Heil's first strategy is his appeal to the *Secondary Identity Thesis*. Recall that an important assumption of Heil's response to zombies was that qualia are just some of the categorical features of the actual world. In various passages across Heil's writing, he seems to take this as a resolution to the problem posed by qualia to his ontology. If Heil is right about this, then it seems that an adequate characterization of mental properties as physical properties is in the offing. He can say that the truth makers for phenomenal truths are dispositional properties of certain systems while resting easy knowing that he is not enumerating properties that are contingently related to phenomenal experience. Similarly with providing an explanation of the qualia in terms of the physical. If Heil is right about the *Secondary Identity Thesis* then he can claim that there are explanations of mental phenomena in terms of dispositional phenomena without being concerned that dispositional properties are contingently related to qualia.

4.5.2 Representationalism

Heil's second strategy appeals broadly to representationalism about phenomenal experience, with a little something extra. Representationalism can be seen as an extension of the idea that at least some of our mental states are intentional – that is, some mental states “represent things as being a certain way”. Generally phenomenal states are considered not entirely characterizable in this way. “Phenomenal states are states with phenomenal or subjective character – something it's like to be in them.” While it might be admitted by most philosophers that there is some representational content to phenomenal states, it seems for many that phenomenal states are not exhaustively representational. Representationalists deny this. Representationalists think that the phenomenal/subjective/what-is-it-likeness of experience is exhausted by that experience's representational content. (Seager and Bourget 2007 p. 271) Since Heil's view differs slightly from this sort of view, I will call the kind of view *Classic Representationalism* and define it like this:

Classic Representationalism The exact meaning of this *exhaustion thesis* is that for every phenomenal character P there is some content C such that a state with P is nothing more than a phenomenal state with C as content. What follows is that, given that a state is in fact a phenomenal state, its phenomenal character is completely specified by its representational content. (Seager and Bourget 2007 p. 273)

To illustrate, consider again Sherrington's reflection on his experience of the sunset. The classic representationalist will answer the apparent mystery that his experience is apparently a physical event in his brain by claiming that the properties that generated the mystery were not properties of the experience itself, but rather properties of the sunset. The classic representationalist will say here that the qualia of visual experience are the “physical properties the brain represents...” the objects we perceive as having. (Dretske 2012 p. 498) In this way, when Sherrington describes his experience of the sunset, he is not really describing features of the perceptual state itself, but rather features of the sunset. (Heil 2012 p. 231)

In an effort to provide *prima facie* support for representationalism Heil brings his readers attention back to U.T Place's famous ‘phenomenological fallacy’. According to this fallacy, we seem incredulous when it comes to the claim that our experiences are events in our brains because we seem to expect the properties of the objects we perceive to be found in the brain itself. Only if we engage in this fallacious form of reasoning do the phenomenal properties of experience take on a mysterious air. (Heil 2012 p. 229-230)

Heil admits, however, that he does not think that representationalism can go all the way in characterizing the phenomenal properties of experience. In this way, Heil is not a classic representationalist since he doesn't think that the phenomenal content of experience is exhausted by its representational content. He defends his claim that there is indeed non-representational content to phenomenal experience in the basis of an appeal to artificial vision, particularly the so-called 'Tactile Visual Stimulation System' (TVSS for short). (Heil 2003 p. 227-230) The TVSS is an apparatus that consists of a camera that is connected to a mechanism that converts the visual image captured by the camera into tactile vibrations on the user's back or stomach. (Heil 2003 p. 228) The device thus enables users to perceive objects at a distance (normally associated with vision) by means of a tactile sort of experience. Heil refers experiments conducted with this device where blind test subjects are able to achieve a significant representational parity with sighted people. (Heil 2003 p. 228-230) Despite this data however, Heil doubts that there will ever be *qualitative* parity between blind TVSS users and sighted people, while admitting that there could in fact be representational parity. He says that whether we see with our eyes or with the TVSS device will make a difference to how the visual experience feels to us. This he claims could happen even if the TVSS technology were to be so advanced that it were able to represent each property of Sherrington's sunset, say, that the human visual system is able to represent.

Heil of course is offering speculation here. He merely is trying to create an intuitively plausible doubt as to whether representational content exhausts the content of visual experience. Whether Heil is able to appeal to the TVSS to demonstrate that there is indeed non-representational content to visual experience is indeed a matter for debate. Classical representationalists, for example, could raise objections at this point. But this is not important in the context of this chapter. What is important here is Heil is admitting, contrary to *Classical Representationalism* that the representational content does not exhaust the content of visual experience. This is not, for Heil, a reason to appeal to distinct mental properties, however. Heil accounts for apparent non-representational content in terms of differences in the *medium* of representation. In the case of a blind person using the TVSS, representation happened through a different medium to the sort of representation that occurs in a sighted person who is making use of their normal visual apparatus. Heil's contention here is that when the medium changes, then so will how the experience *feels*. Sure this is non-representational content, but it is not mysterious according to Heil because it is accounted for by appealing to differences in representational medium. And differences in representational medium could in turn be accounted for by reference to differences in one's physical states – brain states, or states of a prosthetic vision device for example.

Heil seems to take this line because he is trying to reach out to philosophers who are skeptical about representational content exhausting the content of phenomenal experience. He tries here to acknowledge the worry and tell a story about it that would be consistent with an ontology of only physical properties. It is nevertheless still a deflationary measure.

4.5.3 *The Benefits of Classic Representationalism and Heil's Attenuated Representationalism*

What if *Classic Representationalism* or Heil's attenuated representationalism can be defended? It would seem that, we'd have a view of mind that is congenial to one on which there are only physical properties. In the first instance, we have a view on which we can provide physical truth makers for truths about phenomenal experiences. On *Classical Representationalism* the truthmakers for truths about phenomenal experiences will be truths about the objects that are represented in those experiences. In so far as these seem to be uncontroversially physical properties, we have an account of the truth makers for phenomenon experiences that coheres with Heil's ontology of only physical properties. Fred Dretske in a recent paper expresses his enthusiasm for this aspect of representationalism. He writes:

What makes the mind so profoundly mysterious, so spooky, so unlike anything one typically thinks of as physical, is the *representational aspect* of that organ—the brain—causally responsible for our intelligent and purposeful behavior. When nothing you experience, nothing you think, nothing you fear, nothing you want, need exist in the physical world for your thinking, experiencing, fearing, and wanting it to explain—causally explain—why you act the way you do, thinking, experiencing, fearing, and wanting begin to look completely alien. Not part of a scientific picture of the world. A representational theory—assuming it can be successfully grounded—constitutes an elegant rescue from this dilemma. (Dretske 2012 p 498)

If *Classical Representationalism* is right, we can tell a physicalist story about phenomenal experience. All we have is the physical system – say a part of the brain that is responsible for representation – and we have the properties of the objects that we represent.²⁸ (Dretske 2012 . 499) Heil's own version of representationalism is not too different. All it will include in terms of the truth makers for phenomenal properties are truths about representational media, which he claims will be as physically acceptable as the properties of objects represented in experience. In this way, either *Classical Representationalism* or Heil's attenuated version of

²⁸ In cases of bodily sensations and perhaps emotions and moods as well, these latter properties (the properties of objects represented) might include as well properties of human bodies. (Dretske 2012 and Tye 1995)

representationalism provides a solution for Heil's apparent difficulties with eliminativism. Remember the worry there was how Heil's modes of physical substances could not act as the truth makers for claims about the phenomenal character of experience. Whether we accept *Classical Representationalism* or Heil's version of representationalism, we seem to have an account of phenomenal experience as a purely physical phenomenon.

In so far as explanatory gaps pose a problem for being able to explain why it is that two physically divergent items agree in mental attribute, representationalism comes to the rescue again. On representationalism, one can say that qualitative identity consists in identity of representational content. And representational content seems to consist in something other than a system's internal constitution given that it is exhausted by the properties of objects represented. Where it once seemed impossible to explain why two physically diverse items had qualitatively identical experiences, it seems like the mystery is lifted. Martians and humans, while physically different, can have the qualitative experience simply in virtue of being in the same representational state – representing objects that have the same properties.²⁹

Representationalism also provides a solution for so-called explanatory gaps as well for it allows the subjectivity of experience to be explained in terms of physically acceptable representational content. Why is it that a given brain state results in one sort of experience and not another? It is because it is representing one set of properties and not another. So, in so far as explanatory gaps seem to block the ability of accounts like Heil's – which reject universals – to give informative explanations of why two physically divergent systems agree in mental attribute, Heil is able to respond as well to troubles he faces regarding attribute agreement.

4.6 The 'Problem' of Subjectivity and Further Trouble For Non-Reductive Physicalism

In the previous section, my goal was simply to set out the key tenets of Heil's ways of dealing with the phenomenal properties of experience. Each of these ways involved in some sense deflating the notion idea that mental states were inherently subjective in nature. On each of the solutions mentioned above, the subjectivity of experience was explained in terms of

²⁹ My discussion of representationalism so far has limited to representationalism as it regards visual experience. If one were to defend representationalism with the goal of giving a thoroughgoing physicalist account of phenomenal experience, then one would need to give a plausible representational characterization of *all* mental states that involved some kind of phenomena element. This seems to include, but is not limited to bodily sensations like pain, moods, hallucinations and dreams (See Tye 1999).

purely objective properties. If we are to defend this sort of strategy, we must be able to make sense of the idea that the subjective aspects of experience *can* be deflated in this manner. If Heil is to be able to defend such deflation in terms of his *Primary Identity Thesis* and his *Secondary Identity Thesis*, he needs to say why it is that we ought to think that it is plausible that we can count the subjective properties of experience among the categorical properties of the world. The task here is not as simple as claiming that subjectivity is on some sense ‘non-dispositional.’ It seems quite clear that at least part of the subjective nature experience is given in non-dispositional terms. Heil needs to show that this is *all there is* to the nature of subjectivity of experience.³⁰

Similarly, subjectivity may persist as a problem for classic representational accounts of phenomenal experience and Heil’s attenuated representationalism. Few deny that phenomenal experience is in some sense representational, but the idea that phenomenal experience is exhausted by its representational content has been consistently viewed with suspicion. Furthermore, while it is intuitively plausible that the medium of representation contributes somehow to the non-representational content of experience, one might doubt that awareness of the medium of representation of is *all there is* to the subjectivity of experience.

Heil is well aware that worries that the subjectivity of experiences outstrips any physical characterization run deep in the philosophy of mind. So, he tries to meet the worry with a direct response. His response is to argue for a further deflationary measure that is independent of the foregoing strategies. He writes that the apparent distinction between subjective and objective features of experience consists merely in a difference between:

- (a) *A’s undergoing* a particular conscious experience, *C*; and
- (b) *B’s observing A’s undergoing C*, perhaps by observing *A’s* brain. (Heil 20120 p. 245)

For Heil, the subjectivity of conscious experience is not meant to pick out unique properties of experience, but rather to differentiate between experiences individuals are in fact undergoing from third person experiences of those experiences. Heil writes:

Imagine that you are self-consciously visually perceiving a ripe tomato in bright sunlight. You are undergoing a particular experience. Suppose that your undergoing this experience *is* your brain’s being in a particular dynamic state, and that an observer is perceiving your experience by perceiving this state. To keep matters simple, imagine that the observer’s perception is visual as well. These experiences, yours and the observer’s, are altogether different. But this is

³⁰ I will set aside the issue of whether Heil’s *Primary Identity Thesis* is plausible.

not because one is subjective, the other objective, whatever that might mean. As experiences, they are dynamic states of *the individuals* undergoing the experiences, and in that regard ‘subjective’.³¹ (Heil 2012 p. 247)

If Heil is able to establish something like this view about the subjectivity of experience, then it seems that he will have a rather benign conception of the subjectivity of phenomenal experience. On this view of the distinction between subjective and objective, the only difference between the two sorts of states is that subjective experiences are those one is in fact undergoing and objective ones are ones that one is observing someone else undergoing. The only difference between the experiences here is that one is yours with your content, the other is mine with my content – but this shows nothing interesting at all about the difference between subjective and objective experiences.

Whether Heil is *able* to provide a deflationary view about subjectivity is beyond the scope of this paper. Indeed, so is an appropriate analysis of the notion of ‘subjectivity’ that makes it clear how the subjective nature of experience is meant to outstrip the representational content of experience, apparent non-dispositional features of experience and facts about representational media. This issue, I have found, is best left up to one’s intuitions about the character of conscious experience. However, I do want to argue that one does take on a particular commitment if one views the distinction between the subjective and objective as *in principle* unbridgeable assuming an ontology of only physical properties. It is important to notice that the claim here is not merely epistemological in nature. It is not to say that human beings are in some sense limited in our explanatory or cognitive capacities in a way that makes the explanation of subjective aspects of mentality in terms of objective features of the physical world beyond *our* grasp, but perhaps within that of some other better-equipped cognitive agent. (O’Connor 2012) Rather, the failure of explanation runs deeper when one claims that *in principle* the subjective cannot be explained in terms of the objective. (Chalmers 2006)

Philosophers who take this sort of view seem to be committed to an ontology like the one endorsed by the strong property dualists. This is because endorsing a divide between subjective and objective that exists *in principle* becomes more than merely to claim the failure of reduction or to admit pessimism about human explanatory abilities. It is also to admit that we fail to meet a minimal standard for physicalism about higher-level properties. It seems intuitive that for higher-order properties about which one is physicalist, one requires that at least in principle, higher-order property be explainable in terms of the lower order physical

³¹ My emphasis

properties. Just like we are able to explain the hardness of diamonds in terms of their molecular structure, it seems reasonable to expect that we can explain mental properties in terms of physical ones. If this sort of explanation were not in the offing for mental properties *in principle*, then one has strong reasons to doubt the physical acceptability of higher-order mental properties. Moreover, the demand for such an explanation does not seem like one that any physicalist can discharge, or reject as unreasonable. In so far as any physicalist needs to assert some kind of necessary dependence relation between mental states – some of which turn out to be subjective in character – and objective physical states, some sort of expectation that the former can be explained in terms of the latter seems perfectly in order.

With this in mind, I hope to place in the way of the non-reductionist yet another dilemma. Either he admits that there is a subjective element to the character of experience and that the gap between subjective and objective is in principle unbridgeable. In this case, he is not really a physicalist at all. Or he could admit that there is in principle a way to explain the subjective in terms of the objective. The trouble with this latter view is that the non-reductionist will lose any sense in which the distinction between mental and physical is ontologically interesting. Higher-order mental properties on this view, would present no relevant difference with the hardness of diamonds, for example. The only difference may be that mental properties are separated from physical properties by a more complex procedure of explanation.

Conclusion

I have argued that arguments for the traditional views need to assume (P) if they are to establish their conclusions. Each of those arguments appeals either to the success or failure of a form of reduction to establish its conclusion. In so far as (P) is what allows a failure or success of reduction to be of any ontological significance, (P) is essential to these argument. The upshot of a commitments to (P), however was the restriction of philosophical theorizing about the ontology of mental properties. Given (P) we seem committed to views on which mental properties are reducible to physical ones and views on which mental properties are distinct higher-order properties. These turn out to be the only options because the remaining alternative is eliminativism.

Being limited to the traditional view was then shown to be deeply problematic. First because of the apparent unavailability of *Non-Reductive Physicalism* due to difficulties it has accounting for mental causation. Importantly, this particular difficulty of *Non-Reductive Physicalism* arose even with the assumption that the non-reductionist was able to appeal to a supervenience relation between mental and physical. However, in chapter three I showed that appeals to supervenience render *Non-Reductive Physicalism* incoherent. However we characterize the modal strength of the supervenience relation, the relationship between the mental and the physical either is too tight or not tight enough. Independent of worries about the modal strength of the supervenience relation, there was as well the worry that if supervenience is explainable, then it is a reductive relation and if it is inexplicable it is uninformative and consistent with the falsity of physicalism. *Token Identity* was ruled out as a means of disguising *Non-Reductive Physicalism* from its competitors because *Token Identity* is in fact inconsistent with the rejection of type identity.

Limitation to the traditional views was made even more undesirable after I argued that the lines of distinction between *Strong Property Dualism* and *Substance Dualism* are blurred, if not that the former position entails the latter position. If I am right about this, being limited to the traditional views now means that we need to choose between reduction and rather extreme views of physicalism.

I then went on to argue that we need not make this seemingly impossible choice because there is an alternative view of mind – that I associated with John Heil – on which (P) is replaceable. I argued that in so far as this view would avoid the problems of mental causation and zombies, it is a view well worth considering further. These virtues however were presented as contingent upon Heil's ability to avoid eliminativism and explain mental attribute agreement

in the face of physical diversity. While it is obvious that Heil's account ostensibly is such that there is no problem of eliminativism or of explaining attribute agreement, I argued that ultimately this issue would be decided on the basis of whether Heil's view could adequately make sense of the phenomenal nature of experience. I did not, however, make any claims as to whether he is able to establish any of his strategies. This would have been beyond the scope of this paper. However, I did argue that if one took a certain kind of objection to Heil's strategies, a commitment to non-physicalism was again on the cards. I argued that those who doubt Heil's strategies to accommodate the phenomenal nature of experience in the physical world by appealing to an *in principle* distinction between subjective features of phenomenal experience and the objective features of the physical world are really adopting a form of non-physicalist position. In this way, I hoped to place in the way of *Non-Reductive Physicalism* yet another dilemma. If he admits that there is an *in principle* distinction, he is no longer a physicalist. If he admits that there is a way to explain the subjective in terms of the objective, then he no longer seems like a non-reductionist in any interesting sense.

I suggest that the way forward for physicalist minded philosophers to delve deeper into Heil's strategies for accommodating the subjective character of experience into the physical world. Classic representationalism, Heil's attenuated representationalism and his identifying qualia with categorical properties could all be explored further. Since it seems that non-physicalism lies on the other side of an *in principle* objection to the accommodation of the subjectivity of phenomenal experience into the physical world to this sort of view, the physicalist seems to have significant motivation to advocate for one of these three positions. Given the benefits we would gain if one of Heil's strategies were plausible, I think that exploring his view is well worth the effort.

Bibliography:

1. Cameron, R (2008). Truthmakers and Modality. *Synthese* 164 (2):261 – 280.
2. Chalmers, D ‘The Conscious Mind, in Search of a Fundamental Theory’ Oxford 1996
3. Brown, R (2010). Deprioritizing the A Priori Arguments Against Physicalism. *Journal of Consciousness Studies* 17 (3-4):47-69.
4. Dretske, F (2012). Chris Hill's Consciousness. *Philosophical Studies* 161 (3):497-502.
5. Hill. C ‘Locating qualia: do they reside in the brain or in the body and the world? In *New Perspectives on Type Identity: The Mental and the Physical*. Simone Gozzano & Christopher S. Hill (eds.) (2012)
6. Horgan, T ‘From Supervenience to Superdupervenience: Meeting the Demands of a Material World’ (1993)
7. Goff, P:
 - a. *Ghosts and Sparse Properties: Why Physicalists Have More to Fear From Ghosts Than Zombies*. *Philosophy and Phenomenological Research* (2010) 81 (1):119-139
 - b. Philip Goff A Non-Eliminative Understanding of Austere Nominalism. *European Journal of Philosophy* 16 (2008). (1):43–54
8. Heil, J:
 - a. ‘Supervenience Deconstructed’ (1998).
 - b. *From an Ontological Point of View* Oxford (2003).
 - c. *The Universe as we Find it* Oxford (2012)
9. Heil, J and Robb, D *Mental Properties* [*American Philosophical Quarterly*](#) 40 (3):175-196 (2003)
10. Honderich, T (ed.) (2005). *The Oxford Companion to Philosophy*. Oxford University Press.
11. Kim, J:
 - a. ‘Supervenience and Nomological Incommensurables’ (1978),
 - b. ‘Causality, Identity and Supervenience’ (1979),
 - c. ‘Supervenience as a Philosophical Concept’ (1990).
 - d. ‘The Non-reductivist’s Troubles with Mental Causation’ (1993b)
 - e. *Mind in a Physical World* (1998)
 - f. *Physicalism or Something Near Enough* (2005)
 - g. ‘Blocking Causal Drainage, and Other Maintenance Chores with Mental Causation’ (2003).
12. Lowe, E.J ... (2002)

13. Schneider, S Non-Reductive Physicalism Cannot Appeal to Token Identity, *Philosophy and Phenomenological Research* (2011)
14. Searle, J. “Why I am not a Property Dualist” *Journal of Consciousness Studies*, 9, No. 12, 2002, pp. 57–64
15. Thomasson, A ‘Ontological Categories and How to Use Them’ <http://ejap.louisiana.edu/EJAP/1997.spring/thomasson976.html>
16. Seager and Bourget *Representationalism and Consciousness* in ‘A Companion to Consciousness’ Blackwell 2007 p. 261 – 277
17. Tye, M ([1995](#)). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press.
18. Wilson, J ([1999](#)). How Superduper Does a Physicalist Supervenience Need to Be? *Philosophical Quarterly* 50 (1999) (194):33-52.