CAUSATION AND NON-REDUCTIVE PHYSICALISM:
FROM OVERDETERMINATION TO MULTIPLE REALISABILITY

By

THOMAS DAVIES [redacted]

A thesis submitted to the University of Birmingham

for the degree of DOCTOR OF PHILOSOPHY

Department of Philosophy

School of Philosophy, Theology and Religion

College of Arts and Law

University of Birmingham

July 2023

Abstract

The first of three papers tackles the argument from causal overdetermination. I begin with an elaboration of the causal overdetermination argument. Two premises of this argument are focussed upon: the causal closure premise and the overdetermination premise. Causal closure states that for every physical event with a cause, it has a sufficient physical cause. Overdetermination states that there is no widespread or systemic overdetermining of effects, with overdetermination defined as an instance where one effect has two or more causes where any individual cause would be sufficient to bring about the effect. Following this, I provide several assumptions which underwrite those premises: the causal relata assumption, the causal relation assumption, and the causal singularism assumption, which I refer to collectively as "the sparse view". Then, several canonical and original arguments against the causal closure premise are outlined, and the ways in which these arguments involve rejection of the highlighted assumptions is brought to attention. Finally, I consider these assumptions as they pertain to the overdetermination premise. I offer a novel assumption, "notional ubiquitous overdetermination", which supposes a number of sufficient causes across levels of reality for quotidian events. This is defended as an intuitive starting point in opposition to causal singularism, an assumption which prioritises physical causation and which I take to underline the intuitive force of the overdetermination premise.

My next paper concerns the exclusion argument against non-reductive physicalism, due to Kim. I summarise the exclusion argument, which concludes that mental properties are ruled out as causes under non-reductive physicalism under certain seemingly reasonable premises. The paper addresses the exclusion premise specifically, which holds that there

cannot be both a sufficient mental and physical cause for some later event. I consider Edwards' Dictum, holding there to be a tension between vertical determination and horizontal causation, as an intuitive basis for the exclusion principle. The basic idea is that we do not require a prior mental cause, M1, of some future mental effect, M2, where there is already a physical property giving rise to M2 at that time. I consider the basis of the Dictum at length and suggest a disanalogy between the original use of Edwards' Dictum and its use in support of an exclusion principle for mental and physical events. With exclusion undermined, a novel model is presented which allows for autonomous mental causation of macrophysical or everyday events in line with our common-sense understanding. I draw upon work by List and Menzies, and Yablo, to show that mental causes are commensurate with certain effects in a way that microphysical putative causes are not. The other part of my model involves incorporating behaviours as commensurate with mental events such that we can assert that some prior mental cause leads to some behavioural, macrophysical event. With Edwards' Dictum and Kim's exclusion principle undermined earlier in the paper, this model can allow for mental causation at a higher level without excluding base-physical properties as the generative bases of all those mental and behavioural events. The final section of the paper addresses some potential objections to the model. Causation and dependence or generation are thus not in competition, and mental causation is preserved for the non-reductive physicalist.

My third and final paper likewise concerns non-reductive physicalism. Non-reductive physicalism is motivated by multiple realisability, which pushes physicalists away from type-identity given the possibility of multiple physical properties being co-extensive with the same mental property. I suggest that the failure of type-identity, coupled with a persisting

philosophical commitment to the notion that mental properties are dependent on and nothing-over-and-above physical properties, leads to the adoption of a certain type of relation by non-reductive physicalists to characterise the mental-physical relation. I call such relations “thin non-identity”, since they either involve identity somewhere in their formulation- say an identity between causal powers rather than the properties themselves- or they risk dissolution into identity on close inspection. Multiple realisability concerns re-emerge for such positions, if this is right. I suggest that one can satisfy the dependence commitment without maintaining that mental properties are nothing-over-and-above physical properties, and provide reasons for supporting a loosened “Nothing-Left” constraint instead which maintains the intuitive appeal of nothing-over-and-above without overly constraining theory-building. Next, two proposed formulations of non-reductive physicalism are introduced with this loosened constraint in mind, and which construe the mental-physical relation as causal, though not diachronically causal. I show not only that such views, once elaborated, can maintain mental properties are wholly dependent upon physical properties in an intuitively satisfying way, but that such formulations avoid multiple realisability objections and carry various theoretical benefits that prior formulations of non-reductive physicalism lack. Non-reductive physicalism is thus protected from the threat of multiple realisability, and either of my proposed solutions are also shown to fit with the sort of causal solution to the exclusion problem posited in my second paper. My thesis concludes, then, with non-reductive physicalism protected from the overdetermination argument, the exclusion argument, and concerns about multiple realisability, along with novel models of both non-reductive causation and mental-physical relations.

Acknowledgements

There are many people without whom this thesis would never have emerged. I must first thank my supervisors, Scott Sturgeon and Alastair Wilson.

Scott has provided me with invaluable philosophical encouragement and critique in equal measure. He has, throughout the development of this thesis, pushed me toward greater clarity in sometimes-opaque arguments. This demand for the clearest possible argumentation has often helped me to clarify my own thinking. Al, meanwhile, provided me with both fine-grained suggestions and illuminating big-picture thoughts on the direction of my thesis.

I owe a debt to the entirety of the Birmingham Philosophy department, for providing an enriching, challenging, and welcoming community. In particular, I wish to thank Alex Miller-Tate, Kash Sunghuttee, Lucy Prior, Will Sharp, Jess Tait, Kathleen Murphy-Hollies, and Tom Baker.

My close friends outside of academic philosophy, Aaron, Luke, Joel, Rebecca, Rosie, and Chris, deserve enormous thanks for acting both as exceedingly patient sounding boards when I wished to discuss philosophy with those outside the discipline, and for taking me out of the ivory tower when time away and perspective was much-needed.

Finally, my parents Tracey and Peter, my brother Joe, my grandad Bryon, and my girlfriend Aderonke, have all been inexhaustible sources of emotional support and practical advice -

thank you all for having faith in me even when I lacked it in myself. I truly could not have done this without each and every one of you.

Table of Contents

An Introduction To The Thesis And Connections Between The Papers

The papers in this thesis are connected by interest and concern in two elements of physicalist theory; underlying physicalist assumptions, and causation. Indeed, much is said about *causal assumptions* in particular.

This thesis is composed of three papers addressing a number of interconnected issues within contemporary philosophy of mind. My first paper is directly concerned with the argument from causal overdetermination, which aims to secure identity between physical and mental events, reductive physicalism in other words. The aim of this paper *in particular* is to demonstrate ways in which underlying assumptions play a role in the premises of the argument, and in so doing to provide a broad roadmap for engaging with the intuitions which motivate this influential argument. One premise of this argument is that there should not be systemic overdetermination, where an instance of overdetermination is two or more sufficient causes for the same effect. Much time is spent weeding out the ways in which this assumption is rooted in a prior acceptance of physicalism, which the argument seeks to establish. This focus on the assumptions and intuitions motivating physicalists, especially reductive physicalists, runs throughout my papers. If the first paper takes aim at an argument for reductive physicalism, to "make the ground safe" for other views, then my second paper is focussed upon my chosen alternative to reductive physicalism- non-reductive physicalism, which similarly maintains that all properties are in some sense ultimately physical or dependent on physical properties, but which denies strict identity between those properties. I take aim at Edwards' Dictum as an intuitive grounds of the causal view behind the exclusion argument. The exclusion argument, if sound, would

establish that mental properties or events non-identical to physical properties or events cannot be causes of later effects, mental or physical. However, this paper is not purely destructive or critical; once the intuitions behind the exclusion argument are exposed, I develop a new model for non-reductive physicalist causation built upon the rejection of those prior intuitions. In doing so, I hope to "make safe" causation for non-reductive physicalism in particular. My final, and longest, paper turns to non-reductive physicalism itself. While the exclusion problem may be defanged, I argue that non-reductive physicalism faces a different issue relating to multiple realisability. Non-reductive physicalists are often motivated to accept the position, as I was, on the basis of concerns that mental properties could be realised by multiple physical properties- undercutting the possibility of mental-physical identity. I argue that multiple realisability is also a problem, in different ways, for prominent formulations of non-reductive physicalism. By my lights, however, there is a solution, though not one which all non-reductive physicalists may find acceptable. I outline a set of intuitive physicalist commitments, and argue that loosening one of them provides us with philosophical resources to construct a new non-reductive physicalist theory with a sort of causation between physical and mental properties. I also suggest that this proposed theory dovetails well with my causal model for non-reductive physicalists as provided in my second paper.

The thesis, then, is centrally about physicalism, reductive and non-reductive, and the intuitions and causal thinking which shape the debates around mental causation and identity. While each paper is, of course, a stand-alone and independent work, they function as a cohesive whole insofar as I argue against one view in the first paper, defend an alternative in the next, and finally propose a revision to the latter position which saves it

from multiple realisability in the third. I conclude that paper with a discussion of theoretical benefits associated with my suggested solution.

I end my introduction here with a few observations. All these papers do not assume a mere dichotomy or dualism between the physical and mental. Papers 2 and 3 in particular develop theoretical models which fit well to a hierarchal model of reality, where microphysical, chemical, biological, etc. Moreover, the mental is not per-se regarded here as "special" in contrast to the physical, although there is no incompatibility in my work with the notion that the only irreducible level of reality beyond the microphysical should transpire to be the mental. Level-talk warrants some explanation, and is addressed in my second paper, on the exclusion argument. As a gloss, however, one way of thinking about reality is that the properties of physics, say, exist at a different and perhaps more fundamental "level" than mental properties, or properties of biology or chemistry. How these levels are delineated, or whether levels talk is wholly misplaced, is a substantive question and area of work. At any rate, this is all to say that much of what follows, where it relates to causation between levels of reality, or the appropriate model for inter-level relations, might be of interest outside the mental causation and physicalist debates.

These papers, finally, owe much to Jaegwon Kim in particular. His work in clarifying debates for the exclusion argument, for non-reductive physicalism, the causal overdetermination argument, and problems surrounding the metaphysics of token-physicalism, as well as work on the metaphysical underpinnings of all these debates, has been vital, and "footnotes to Kim" could well be thought of as a further thematic thread connecting the following papers.

Paper 1: Causation and "The Physical" in the Argument from Causal Overdetermination

1.1 Introduction: The Overdetermination Argument and the Sparse View

The Causal Overdetermination argument is an argument for reductive physicalism, and against any interactionalist dualist account of the mind-body relationship[1] (Gibb, 2015.c, p.70/71). It is intended to establish that mental events are identical to *some* physical events, and runs thusly:

1. **Mental Causation**: 'At least some mental events are causes of physical events' (Lowe, 2000, p.572)

One might suppose that there could be some totally non-causal mental occurrences or events, but this is not particularly important to this argument, nor its conclusion. Kim, for example, is happy to accept that there might be certain "free-floating" non-physical properties or events (Kim, 2007, p.13), but this argument is concerned with events which interact in the causal-physical domain.

2. **Causal Closure**: 'Every physical event which has a cause, has a sufficient physical cause' (Lowe, 2000, p.575).

[1] Unlike epiphenomenalists, any sort of interactionist dualist Is committed to the first premise, which follows.

There are many ways in which this premise is formulated within the literature, discussed in a later section. A toy example may help to illuminate what such principles are seeking to pick out, however.

Take some caused event, such as a mug being picked up. There is some physical cause, and causal story, explaining how that mug came to be picked up. This will involve bodily movements and neural impulses at the biological level, and patterns of microphysical interactions below that.

3. **Overdetermination**: There is no systematic overdetermination (Papineau, 2001, p.9).

An event would be overdetermined just in case it had two or more causes, all of which were individually sufficient to bring about precisely the same event (Crane, 1995, p.19). One canonical example is the case of a man being shot, at the same time as he is struck by lightning. His death is *overdetermined*. To be clear, the "Overdetermination" principle, is not claiming that cases of overdetermination are impossible, or incoherent. The claim is merely that they are not systematic: we do not observe or expect events to be overdetermined on a regular basis.

Elaborating further, when I refer to the ruling out of *systematic* overdetermination, I am attempting to capture a sentiment common across formulations of overdetermination principles (Sturgeon, 2000, p.123, Papineau, 2001); that events are 'not generally

overdetermined (Sturgeon, 2000, p.123), and that overdetermination is a weird coincidence when it does occur, such that it should not be expected as a regularity.

Conclusion: 'Mental events (that are causally relevant in the physical domain) are identical to certain physical events' (Gibb, 2015.c, p.70).

The premises of this argument are discussed in greater detail across following sections. For now, I will provide some brief, though necessary, elaboration. To make this argument valid, either the first or third premise requires modification. If the Mental Causation premise does not stipulate that the mental causes are also sufficient causes, then there is no tension with Overdetermination which *only* denies the ubiquity of sufficient causes, as opposed to causes *simpliciter*. One may either stipulate, regarding the Mental Causation premise, that there are at least some *sufficient* mental causes of physical events, or else broaden the Overdetermination premise such that it rules out not only two or more sufficient causes but two or more causes of any kind for a given effect.

Mental Causation, for those who endorse it, is taken to be relatively ubiquitous, and a regular feature of our world. Mental causation occurs when I write these words, when my reader moves their eyes down the page, or scrolls to the next, when one moves to make the morning coffee, and turns the lights off before bed. Simply put, it happens all the time. Therefore, it would be unreasonable for one endorsing the first premise to evade premise 3 through an appeal to some notion that mental causation is rare or occasional. It is also plainly unsatisfactory to anyone holding Mental Causation. With this aside, the conclusion

seems to follow from the premises[2]. Given the lack of systematic overdetermination, and the implication that any given caused physical event must have some physical cause, the only way to save the first premise, mental causation, is to conclude that the mental and physical causes are identical and, consequently, that physically causal mental events are identical with certain physical events.

To elaborate, we can imagine some set of physical events such as "all instances of a person picking up a pen after making the conscious decision to do so". Each instance in the set has a mental cause, per Mental Causation. From Closure, since each event in the set has a cause, they must have a sufficient physical cause. Call this set of physical causes P[x], the set of all corresponding mental causes M[x][3], and discrete physical and mental causes within the sets P and M respectively. Take some physical cause P of the token physical event, and some cause M as the mental cause. If P and M are distinct causes of the event, then the event is overdetermined. If this is the case for *every* instance of these caused events, and corresponding sets P[x] and M[x], then we have regular, indeed frequent, overdetermination. However, this is ruled out by "Overdetermination". Since P has a sufficient physical cause, and a mental cause, and we do not wish to claim that P is overdetermined, we ought to conclude that M is identical with P1.

The causal overdetermination argument does not entail, necessarily, that all mental properties (uninstantiated properties, for instance), must be physical (Crane, 1995, p.9). Nor does it entail *all* mental events are physical, only those that are causally relevant in the

[2] In what follows, however, I intend to suggest that this argument depends on certain implicit assumptions.
[3] The conscious decision to pick up the pen in each case.

physical domain; if the argument is sound, then all mental events which are causes of physical events are themselves physical events.

This argument takes events to be the causal relata (Hall, 2004, p.1, Gibb, 2010, p.365). Following Papineau (2001) and Kim, I will take an event to be the instantiation of a property, in a substance, at a time (Gibb, 2010, p.365). This account of events is neutral on whether the properties, or the substances in which they are instantiated, are physical or not. Event causation does not *prima facie* rule out mental events, or any sort of non-physical event. It is worth noting here that many within this debate talk about causal sufficiency, with regard to physical and mental events.

An event is causally sufficient for another just when its occurrence is sufficient, or *enough*, to generate the effect. Mackie introduced the notion that, when we talk of the sufficient cause of an event, our meaning is more accurately described by what he termed an INUS condition. This refers to 'an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result' (Mackie, 1965, p.245). Mackie's example of an INUS condition is a short-circuit in a house as a *cause* of a fire (p.246). Here, the short-circuit was insufficient (on its own, without the presence of flammable material nearby, and the lack of a sprinkler system and so forth) to start a house fire but was necessary, in this circumstance, since it provided the ignition source. This *confluence* of events was not the only way a house fire could have started (arson or a lightening strike would be alternatives), hence it was an *unnecessary* condition, but clearly *sufficient*, in light of the fire. This confluence is the INUS condition. Building on INUS, Paul and Hall developed the Minimal Sufficiency account of causation:

'C is a cause of E iff C belongs to a set of contemporaneous events that is minimally sufficient for E; i.e., sufficient, but with no proper subset that is sufficient. And a set S of events occurring at some time t suffices for a (later) event E iff, were the events in S the only events occurring at t, E would (still) occur' (Hall and Paul, 2013, p.72).

Taking physicalism to be the claim that all 'properties are realised by, or identical to, physical properties' (Gillett, 2002, p.92), or even simply that all entities and properties are physical, the Causal Overdetermination argument does not establish physicalism *simpliciter*, given the aforementioned possibility of non-instantiated properties. It does, however, rule out any form of interactionist dualism, understood as any position positing 'independent' (Montero, 2003, p.178) mental causes of physical events. As Gibb suggests, interactionist forms of both substance, and property, dualism are threatened by the argument, since both posit mental events as the instantiations of irreducibly mental properties, which have causal impact on the physical (Gibb, 2015.b. p.132).

This paper will leave the mental causation premise untouched. There are those who wish to deny that mental events, properties, particulars, or states of affairs, have any causal impact in the world. Those who hold such views are not the target of my forthcoming critique, though they may also find compelling reasons within this paper to reject (2), (3) and the assumptions I take to underly them, even if nothing about their view of mental properties is affected.

My concern, presently, is with Causal Closure and Overdetermination. There are many reasons to be concerned with both premises, to be discussed. However, I do not take these issues to be isolated from one another. A central contention of this paper will be that both premises are rooted in a particular set of inter-related assumptions which need not be accepted, and moreover, that one of these assumptions- absent independent argument- must presume the reductive or identity-physicalist conclusion which the overdetermination argument is seeking to establish. As these assumptions are introduced, I will demonstrate how the rejection of any one allows one to undermine the Overdetermination argument.

Following the next section, I canvas a number of canonical and original arguments against the overdetermination argument, and my early sections will focus on the closure premise in particular. These arguments do stand alone against closure, but I hope to show how the assumptions I elaborate henceforth are rejected by these arguments. We can then note how these arguments are functionally rooted in the rejection of one or more of the assumptions.

Now, I will elaborate on the assumptions themselves.

I term this cluster of assumptions “the Sparse View” for reasons explained shortly, and term those physicalists who accept them, explicitly or implicitly, sparse physicalists. I identify four; the “causal relata” assumption, the “causal relation” assumption, and the “causal

singularism" assumption[4]. As noted, each will receive significant elaboration, though I outline them briefly here for clarity and reference:

(A) Causal Relata: All causal relata are members of the same metaphysical class or category.

(B) Causal Relation: There is only one sort of causal relation, or one way in which a mental particular may *be* a cause, or causal.

(C) Causal Singularism: Causation is constrained to one level of reality.

As will become clear within arguments offered later in this paper, many objections to premises 2 and 3 already explicitly or implicitly deny one or more of these assumptions. It is important, nevertheless, to outline and examine these assumptions to develop a better understanding for where the argument goes wrong, and to get a grasp of the intuitive underpinnings which make the argument so *prima-facie* compelling.

As I see it, the causal singularism assumption broadly explains why the sparse view "hangs together", so I will begin by elaborating this assumption in detail, and how its rejection undermines the third premise of the overdetermination argument. Causal singularism is a pre-argumentative commitment to a singular "causal" level of reality conjoined with an intuitive picture of causal relations "flatly" working at that level. I do not intend to argue

[4] Edwards' Dictum, a principle which holds there to be 'there is a tension between vertical determination and horizontal causation' (Kim, 2003, p.153) may warrant inclusion on this list, at least if we were merely canvassing assumptions underlying (3). I discuss the Dictum at length in another paper and will therefore not dwell on it here.

that this “picture” is wrong, but that it does too much work in establishing the intuitive force of (3) to go unexamined or unsupported with independent argument.

To elucidate more clearly what this means, I will introduce a contrasting starting-point assumption; notional ubiquitous overdetermination (henceforth NUO). NUO holds that there are systematically multiple causes for many quotidian events. I take this to be supported by the prima facie observation that, for any everyday event, there are multiple apparent causes across different levels of reality[5], or at least from different fields of science and folk-psychology.

To see why one might observe this, consider our practices of causal explanation in one toy example. John walks into a shop and steals some valuable items, and we want to know what caused him to do this. A reasonably well-informed person might be able to proffer, or at least reference, a host of non-competing, non-exclusive, explanations, a selection of which I now provide.

First, there is a social or folk-psychological explanation pertaining to John’s financial situation, his tolerance for risk-taking, his upbringing and so forth. John is made desperate given the state of his finances, he tends to solve his problems without consideration of negative consequences, and the shop provides a seemingly easy target. His desire to escape

[5] Some philosophers, like List (2019, 2022) take levels of reality to be ontologically real and indexed to at least certain levels of scientific description and explanation. Even Kim, a proponent of the overdetermination argument, cautions that the metaphysics of levels talk warrants being taken ‘seriously’ (2002, p.3). I do not intend to provide a defence of such a view here, only to highlight that contrary views do not warrant presumptive acceptance for purposes of the overdetermination argument. My reasons for this are elaborated shortly.

his current situation, his disregard of aforementioned negative consequences like imprisonment, and his belief that he can achieve his goal through stealing from the store (in short, his mental state prior to the theft), generate a volition which causes him to steal. A psychiatric causal explanation may also be provided, drawing on some of the same factors but incorporating as elements some disorder which in particular causes John to act as he does. I omit further detail for sake of brevity, but one may deny that this is an explanation independent of the folk-psychological causal explanation. Consider, then, a fully biological or neuroscientific explanation, as follows. John experiences synaptic activity and dopamine release pertaining to rewarding stimuli, the high-value items in the shop, he experiences hunger impulses due to lacking ability to provide for himself, and so forth. Finally, and more fine-grainedly, neuroscience will point to a chain of neural activity which eventually causes John to walk into the shop and steal. This, too, seems a complete and self-contained causal explanation for John's action.

Even besides these causal explanations, we might consider still more. Explanation, step by step to John's theft, of the chemical interactions within John which caused the crime. Even "below" this, we can find a microphysical causal explanation telegraphing the movement and interaction of the set of John-shaped, and other relevant, particles toward and through his theft.

Some of these explanations may seem situationally inappropriate; one is not likely to offer a chemical or microphysical story to explain why John stole, but it is doubtful that any would deny that such a causal explanation fits. One may further suppose, pre-theoretically, that such causal explanations, at least sometimes, provide pro-tanto evidence for real causes

and causal relations (Skow, 2014). Given this, we likewise have pro-tanto reasons to attribute a host of causes to the same event: "John stealing from the shop". I think most everyday events can likewise be provided with such a list of causal explanations, and hence, *notional* causes. Put simply then, at face-value, these layered facts in the world provide us with commitment to notional overdetermination. I state "notional" here since this remains assumptive; we seem to have this array of causes in light of this observation, even if this ultimately transpires to be incorrect. Nevertheless, the possibility warrants consideration since, if right, overdetermination goes from a particularly peculiar, at best uncommon occurrence, to a pervasive phenomenon[6]. Further, it presents us with a picture of the world where intra- and inter-level causation are common, and does not privilege any particular level of causal explanation or causation. This NUO picture thus stands in contrast to causal singularism.

One objection to NUO is worth noting here, even though I do not present NUO as an argument for real systemic overdetermination so much as a presumptive starting-point for metaphysical consideration, in contrast to causal singularism. The objection I have in mind, however, would seek to undercut the notion that there really is even *notional* overdetermination, and runs as follows:

(1) "Higher level" explanations, or explanations from scientific fields outside of physics, are targeting the same causes and physical phenomena, merely at different levels of complexity and abstraction.

[6] Other independent argumentation for overdetermination is offered later in the paper- NUO is only discussed now to shed light on causal singularism.

(2) Causal explanations pick out real causes, but given (1) we should not expect them to identify these causes as microphysical.

(3) All concrete entities picked out by higher-level causal explanations ultimately decompose into microphysical entities.

(C) Causal explanations at levels besides physics are merely using different predicates to pick out base-microphysical causes.

I take this, and any other appeal to decomposition, or reduction, to beg the question against NUO within the context of the overdetermination argument. The argument sets out to establish the identity of certain higher-level, mental properties, with microphysical properties. One should not, then, appeal to an assumption of identity in support of the intuition behind one of its premises, or against the contrary intuition.

A related concern regarding NUO is that the suggested causal explanations in the theft case may merely pick out something like causally relevant conditions for the titular event. "Causally relevant" here is only intended to indicate that these events were not individually sufficient to bring about the event, even if they seem non-exclusively causally contributory.

In response, it must be first be reiterated that NUO is a *starting-point assumption*, not an assertion that there certainly is causation across many levels of reality or within many fields which stipulate causes. The thought, to elaborate, is that our everyday *and* scientific practice of causal explanations give reason to presume causation across these levels, shifting the burden of proof to those wishing to deny such. It is also hoped that NUO serves

to take some of the intuitive "sting" out of systemic or ubiquitous overdetermination, often framed as an inherently *peculiar* possibility. One may accept NUO and remain open to the possibility that causation is misattributed, or somehow derivative, outside physics. However, *if* NUO, the assumption of causation across levels and outside of microphysics, is a reasonable starting-point assumption, then it is surely due to its fidelity to our experience with causal attribution across previously noted contexts. We would not, intuitively, assert that the complete folk-psychological explanation of John's action were merely causally relevant; we would say they caused him to act as he did. Likewise, we would not say a sufficiently complete biological explanation presented events which were merely causally relevant; we would attest that they caused his actions, and so forth. To convincingly argue for substituting causation for causal relevance in these cases, it seems to me that one owes an explanation for how we are misunderstanding everyday and scientific causal talk which actually does not designate causes, or how these intuitions themselves are somehow faulty or misleading. In the absence of such explanations, NUO remains intact as a viable starting-point assumption.

Besides this, one could also adopt the metaphilosophical stance against deploying intuitions for this purpose; this would be an ad-hoc move in isolation, however, and would also fail undermine the overdetermination premise independently, insofar as it relies upon a competing assumption with competing intuitions underlying it- causal singularism. It is that assumption which I shall now address.

With NUO explained, and briefly defended as a plausible assumption, we can consider causal singularism and what it entails. At the most basic level, I take causal singularism to be

the rejection of, or inverse of, NUO[7]. Once again, this rejection is not normally an explicit one, but takes the form of a presupposition that causation is restricted to the most fundamental level, and the pro-tanto disregard of causation at higher levels. Caveating this, however, Kim is one philosopher who is more explicit in a rejection of NUO and acceptance of causal singularism as a starting point:

'Moreover, just saying that there "obviously" are biological causation, physiological causation, and so on isn't very helpful; what has to be shown is that these kinds of "higher-level" causation are irreducible to basic physical causation—namely, that there are these causal relations in addition to the underlying physical causal processes' (Kim, 2005, p.54).

Here, Kim takes it that the burden of proof lies upon anyone arguing for causation at higher levels, while taking for granted that the 'underlying physical [] processes' (ibid) are causal.

On causal singularism as a starting-point assumption, overdetermination appears rare at best, and systemic overdetermination seems extremely implausible. This is not just because notional overdetermination is brushed aside, but because of consequent influence on our

[7] It should be noted that NUO can be rejected without endorsing causal singularism, and as such causal singularism should not be understood as equivalent to the rejection of NUO. For example, philosophers could endorse a "parallelist" ontological view which holds that intra-level causation occurs between same-level causal relata, without accepting the occurrence (or perhaps even metaphysical possibility) of inter-level causation. This causal parallelism would differ from NUO in this rejection of inter-level causation, *and* from causal singularism in positing levels besides fundamental physical reality. As a prospective starting-point assumption, I do not readily see the intuitive pull that would motivate one to adopt some sort of parallelism as the most pre-theoretically plausible option. We can remain agnostic on parallelism and bracket it for purposes of this paper for two reasons. First, the parallelism would be incompatible with the principle of mental causation inasmuch as it denies any inter-level causation, including between mental and physical levels. Given that the focus of this paper is upon those who are committed to the principle of mental causation, this incompatibility would seem to rule out parallelism as a plausible candidate for these philosophers. Secondly and relatedly, my concern in spelling out NUO is not to fill the entire concept space, but only to provide one reasonably intuitive *alternative* starting-point assumption to causal singularism while also helping illuminate the concept and implications of singularism.

intuitive causal picture. Since causal singularism constrains causal relations to one level of reality, when paired with a reasonable commitment to causal relations within the microphysical domain[8], microphysical causation acquires an apparent privileged status. Therefore, in considering how causal singularism generates the other sparse assumptions, we should consider how philosophers tend to intuitively characterise or think about microphysical causation, and causation within physics more broadly. Consider the following;

'The astonishing accomplishments of fundamental physics raise the possibility that our world contains relatively few kinds of fundamental microscopic entities and quantities that, together with its spacetime structure and a few fundamental laws, are sufficient to account for the motions of all material bodies and every other event that physics expanded to include' (Loewer, 2001, p. 38)

Loewer's account is instructive here. To summarise, we have microscopic entities (and quantities thereof), structures formed from these entities, and laws governing the causal interaction between these basic entities. Bracketing, for a moment, concerns about the status of causation at the lowest level of microphysics on certain theories[9], we are left with a picture of basic entities organised or composed in certain ways and entering into same-level relations. Modern fundamental physics takes there to be two sorts of subatomic particles which interact via the four fundamental forces; fermions have mass and compose

[8] One need not even accept any formulation of the causal closure principle to accept that there are causes and causal relations between microphysical entities. Certainly, I do not intend to deny such causes. Parenthetically, some philosophers are concerned that causation might not be a microphysical phenomenon, but instead a feature of the macrophysical world from the level of thermodynamics and up (Ney, 2016, p.146).

[9] For more on this possibility, see Campbell & Bickhard (2011), and Ney (2016). I also discuss this in my second paper, as it relates to causal closure within the Exclusion argument.

protons and neutrons, while different kinds of bosons mediate the four forces themselves (Schwichtenberg, 2018, p.7-8). Though mathematically complex, the overall causal picture is a simple one; a limited set of fundamental entities being held together by the strong nuclear force, repelled or attracted by the electromagnetic force, and having their electric charge altered by the weak force (ibid, p.169, p.173). No evidence for systemic overdetermination is apparent within these discrete interactions. One worry, however, is that this may simply be because there is no causation within fundamental physics at all.

The existence of causation at the level of fundamental physics is not a topic I intend to delve into at length, since this would warrant its own thesis-length treatment. However, I will now provide a few brief remarks in relation to that debate before moving forward, insofar as it is relevant to the overdetermination argument and causal singularism. First, *if* the physical causation picked out by Closure is not found at the level of fundamental physics because there is no causation at that level, then we must determine what level proponents of Closure have in mind. Ney notes this tension, between the Russellian eliminativist view on microphysical causation- 'the view that causal notions have no place in microphysics, even if they may correctly be used to describe nonfundamental facts in the folk or special sciences' (Ney, 2016, p.142)- and the appeal to microphysical causation within various formulations of the Closure principle (Kim, 2005, p.53-54, Lowe, 2000) discussed in a later section. Opposition to microphysical causation, originating with Russell, is typically rooted in models of causation which characterise it as a time-asymmetrical relation, contrasting with 'fundamental physics [which] describes the world in terms of dynamical laws that are, possible small exceptions aside, time symmetric and that relate global time slices' (Loew, 2017, p.1945-1946).

One can attempt to defend microphysical relations as causal through abandonment of time-asymmetry as a disqualifier on causation, or through the identification of specific quantum phenomena which can, at least, be formalized time-asymmetrically (Adlam, 2022). One may go further still, and provide some argument that time-asymmetric causation is a common feature of sub-atomic or quantum interaction (Bartels and Wohlfarth, 2015).

We might adopt a different approach, following Papineau (2013), and suggest that the relevant causal level of physics is macrophysical, thermodynamics (ibid). Laws of thermodynamics are time-asymmetric, given 'that entropy is very likely to increase toward the future' (Ney, 2016, p.146) as per the second law of thermodynamics. The time-asymmetry within the second law is not typically understood to be reducible to microphysics alone, but to the fundamental microphysics when conjoined with 'a probability distribution over the microstates capable of realizing a given macrostate' (ibid, p.147) and the initial conditions of the universe at low-entropy (ibid).

Whether one finds a place for causation within fundamental physics, or in thermodynamics, the *relevant* intuitions resultant from causal singularism hold, as I will now explain. If one adopts causation as a macrophysical, thermodynamic, phenomenon following Papineau, then one still takes causation to occur between events; particle interactions governed by the four laws of thermodynamics. The Closure-endorsing causal singularist still emerges with causation at a level of reality which is functionally mechanistic, with interactions between discrete nodes governed by a limited and intuitively simple set of laws. Further, thermodynamics is directly and explicitly concerned with heat, or energy, so the philosopher

taking thermodynamics to be the level at which causation pre-theoretically occurs has clear reason for adopting a metaphysics of causation which centres energy transfer or flow of momentum. It is straightforward, then, to see how causal singularism drives philosophers toward 'energy flow or momentum transfer' views of causation (Kim, 2007, p.17) given the manner of causation observed at this level of physics. One need not, of course, adopt such a view of causation even if one were to explicitly accept causal singularism, yet it seems to me that the intuitive force of such views on causation flow out of the implicit acceptance of singularism.

However, these options may risk missing the forest for the trees, so far as causal singularism is concerned, and given its status as an implicit assumption. The serious metaphysical questions of whether causation is necessarily time-asymmetric, and whether causation is to be found within fundamental physics seem in no sense philosophically pre-theoretic. If one disagrees on that point, I hope my foregoing discussion serves to resolve any concerns, regardless. Such disagreement aside, and taking causal singularism seriously as a pre-theoretic assumption, I think we can reasonably assert that the singularist picture one way or another will principally involve a circumscribed set of particulars under similarly limited rules, and without any special forces or peculiar exceptions. Papineau articulates this intuitive physicalist picture of the base-causal physical world while providing his own argument for physicalism, which I discuss further in Sections 3.B and 3.C:

'All apparently special forces characteristically reduce to a small stock of basic physical forces which conserve energy. Causes of macroscopic accelerations standardly turn out to

be composed out of a few fundamental physical forces which operate throughout nature' (Papineau, 2001, p.19)

Further, and focused upon fundamental physics, we have Smart:

'Certainly we are pretty sure in the future to come across new ultimate laws of a novel type, but I expect them to relate simple constituents: for example, whatever ultimate particles are then in vogue' (Smart, 1959, p.143)

Whether fundamental physics or thermodynamics, causation at the most basic causal level of physics appears *simple* insofar as we are concerned with the number of posited kinds of basic particular, and the small set of laws or forces determining their interactions. This apparent simplicity bears on the following two assumptions, and other aspects of the debate, as I will discuss shortly. Causal singularism and NUO themselves will be discussed further in Section 3, where I shall consider overdetermination principles more broadly and the ways in which causal singularism might lead us wrong.

For now, however, we can turn to the other assumptions of the sparse view. While one can hold "causal relata" or "causal relation" independent of the causal singularism assumption, I should first note *why* I take the two following assumptions to follow if one already accepts singularism. "Follow" in this instance does not denote logical or conceptual necessity, but rather means that singularism generates strong intuitive appeal for the other assumptions, and conversely that non-endorsement of singularism undercuts their intuitive appeal. If we take ourselves to inhabit the sort of flat causal landscape implied by singularism, then it is

straightforward to accept that all the causal relata are of the same metaphysical, and perhaps physical, type, since we are only intuitively open to these fundamental, *simple*, particles as causal entities. Further, our stance on causal relations is informed, in singularism, only by observation of physics with its circumscribed set of causal interactions[10]. The "causal relation" assumption holds that besides the nature of the causal relata, there is only one sort of causation or causal relation itself, at least insofar as concerns the mental and physical[11]. This assumption seems central to the argument as presented across various canonical formulations (Crane, 1995, p.4, Papineau, 2001). Typically, the sort of relation in question is event-causation (Lowe, 2000, p.572). Event-causation, rather than type-causation or fact-causation, seems to fit well with the Sparse View assumptions which jointly centre physical particulars as the locus of metaphysical inquiry. Indeed, to the extent that the causal relation assumption has intuitive appeal, we might suppose this is in virtue of the apparent simplicity of causal interaction within physics, wherein discrete and individuable particles, or sub-atomic particles interact, per the singularism assumption.

Both the mental causation and closure premises refer to events, except where the argument is framed in probabilistic terms or otherwise to avoid metaphysical commitments. Nevertheless, even in less metaphysically committal framings of the closure principle which refer to physical causes merely determining or fixing the chances of some later effect, as I

[10] Charge, spin, and momentum as governed by the fundamental forces, to name salient examples.
[11] One could ask, quite reasonably, why I do not identify this assumption as causal monism, given that position already exists in the literature (Wahlberg, 2022, p.3276). My assumption extends beyond what causal monism asserts. Causal monism holds that there is only one sort of metaphysical causal relation. I think one who holds the "causal relation" assumption will accept this, but I also take them to be committed to homogeneity of causation in other forms, addressed shortly.

discuss in Section 2, the same sort of determining or fixing is posited for the mental cause. If this assumption is undermined, then one promising route to establishing a closure principle in the first place is cut off, for reasons elaborated on in Section 4.

The "causal relata" assumption holds that all causal relata (or mind-body relevant causal relata, at least) belong to the same class of metaphysical entity, whether properties, events, states of affairs, or facts[12]. Per singularism, we are also provided with a view of reality where these relata are likely individuated in a fine-grained manner, as with microphysical particulars implicated in the causal explanations of physics. If this assumption is not accepted, then the door is open to consideration of distinct relata in cases of mental-physical causation. I elaborate in 1.2.3.1 but briefly, consider that there might be some microphysical chain of causation between events, while some distinct mental cause is instead taken to be the cause of some other, higher level physical effect of a different metaphysical kind. This steps beside, in effect, any sort of overdetermination premise, while preserving microphysical causal closure and permitting mental-physical causal efficacy.

It must be noted that the causal relata and relation assumptions are likely to stand or fall together in particular. I offer one way this might occur later in the paper but to introduce the notion we might consider that one way for something to be involved in a different sort of causal relation is through being a *different kind* of cause.

[12] Though I take this, evidently, to extend to reductionist positions which do not hold events as the causal relata, I continue to use "events" through much of this paper for the sake of simplicity. In those places where the nature of the relata becomes relevant, it will be discussed explicitly.

The rejection of any of the "sparse" assumptions (A)-(C) is sufficient to undermine the force of the overdetermination argument. Some of these reasons have been discussed already, while others will be covered as this paper considers the relation of these assumptions to certain arguments against premises (2) and (3).

I term this set of assumptions the sparse view, since I take the overall cluster to denote minimal ontological commitments, or commitments to only a sparse set of causal phenomena. Besides this, there is an instructive comparison to be made with *sparse ontology*;

> 'Central to sparse ontology is a hierarchical view of reality; the basic entities form the sparse structure of being, while the derivative entities form the abundant superstructure'
>
> (Trogdon, 2008, p.147).

On sparse ontological views, we can intelligibly talk about the derivative entities which are involved much of our everyday experience and scientific practice, but they are not to be understood as anything "over-and-above"[13] structures of these basic, fundamental entities. Centrally, there is no ontological commitment to these derivatives over and above the commitments we hold to these basic entities (p.153).

For a similar summary of a view which minimizes ontological commitments, here is "cheap truth theory";

[13] Sparse ontology demands nothing-over-and-aboveness (Trogdon, 2008, p.153) of any non-fundamental, derivative entities. "Nothing over and above" is a term of art which I address at length in another paper.

‘Cheap truth theory is not just the view that simple concrete entities are metaphysically basic, but that they are the only entities that we are ontologically committed to.’ (Goff, 2016, p.895).

The sparse view seems to offer a likewise “cheap” account of causation. It postulates only *one* sort of causal relata, typically events, causally interacting in *one* manner, likely some sort of energy or momentum transfer, at *only* the most basic level of microphysical reality. On this view, one could map causes and effects on a 2-dimensional grid with only arrows denoting causation, and lose little content regarding the causal relations at work. It is “flat”, it posits few sorts of metaphysical entity, and it is an intuitively simple picture.

The assumptions of the sparse view could be well-founded, or correct, although I will offer several reasons to think otherwise. For now, however, it must be noted that if these assumptions underwrite the premises of the causal overdetermination argument, they must be supported for reasons which are independent of the reductionist or mind-brain identity theories which the overdetermination argument seeks to ground. Absent such independent reasons, the argument will remain unconvincing to the non-reductionist who need not accept these assumptions.

It is important to note that my sparse view should not be confused with Wahlberg’s sparse causation (henceforth SC), defined here:

‘SC: A token instance of SC consists in some appropriate physical process connecting worldly entities’ (Wahlberg, 2022, p.3261)

There are commonalities; like the sparse view, SC focusses on physical causation, and “appropriate physical process” seems to limit the manner of causal relation much as the causal relation assumption. However, sparse causation represents one form of causation within Wahlberg’s pluralist account (2022, p.3277); it is embedded within a tripartite causal framework (ibid), rather than being positioned as the *only* sort of causation, or a cluster of causal assumptions such as the sparse view.

The assumptions of the Sparse view yield minimal ontological commitments, and a likewise minimalist picture of causation in the world. In what follows I aim to show, both through my discussion of these assumptions, and arguments against the explicit premises of the causal overdetermination argument, that this picture is in fact be *too* minimal. My intent in providing many of these arguments, some of which are my own and some of which have been developed by others, is not to merely link them to these assumptions,. Rather, I will demonstrate that one can, in many cases, develop plausible accounts of causal relations between the mental and physical, starting on the basis of a rejection or reversal of one of more of the sparse assumptions. In short, we need not reject these assumptions without viable alternatives.

The ”sparse physicalist” is the only philosopher whose starting-point assumptions compel them to accept the premises of the argument from causal overdetermination. Indeed, I will

argue that even physicalists need not be a *sparse* physicalists; and perhaps should not wish to be.

Before moving on from the sparse view, I should emphasise again that these clustered assumptions constitute a pre-theoretic, or intuitive, picture of causation, which underline the intuitive force of the causal overdetermination argument. The sparse view, then, is not to be understood as a full theory of causation[14] itself, even if it might rule out certain theories and provide intuitive support to others. If one undercuts these motivations for the premises, then the argument from overdetermination itself is likewise undermined. Work on underlying assumptions is more readily apparent in work on causal closure, due to Gibb whom I discuss shortly, and so my references to the assumptions of the sparse view will be less centred, as the paper concentrates upon merely identifying assumptions already picked out. My final section more concretely centers the Sparse View as I consider the assumptions at play in support of overdetermination, which I do not take to be as well-explored in the literature- at least, not in the terms I am explicating.

With the causal overdetermination argument and the sparse assumptions laid out, my paper will run as follows. The next section considers the second premise, Closure, in greater detail. I canvas the "standard" options, besides rejection of Closure, available to those who wish to resist the conclusion of the argument. I then address some tacit metaphysical presumptions of the argument.

[14] By full theories of causation, I mean theories which claim to give a complete metaphysical account of causation, such as counterfactual theories, productive theories, regularity theories, and so forth.

The second section will examine various formulations of Closure put forth in the literature, delineating stronger and weaker claims, as well as certain forms of Closure which are probabilistic, or causal, respectively. After offering a taxonomy of Closure formulations, I will settle on one formulation to be under consideration for the rest of the chapter.

In the third section, I will introduce two forms of argument aimed at motivating Closure. Firstly, there are a-priori defences which insist that the causal closure of the physical domain is a starting assumption of science, or even entailed by the *in-principle* completeness of physical explanations. Secondly, there are arguments due to Papineau, which point to the success of reduction to fundamental forces within physics, and to the failure to find so-called special forces, in neurophysiology. I will offer reasons to suppose that a-priori defences of Closure are, variously, question-begging, and unsupported in their historical and predictive claims. Moving to success-of-science arguments, I will argue that there is a clear lack of evidence for the underlying assumptions made by Papineau. I will return to the argument from physiology in particular, later in the chapter, when considering the role of explanatory adequacy in claims about causal closure.

Before this, however, my fourth section will contend that the conjunction of Papineau's conclusions, even granting them a higher level of credence than I have argued is warranted, still fails to entail Closure. I will argue that more is needed for a valid Closure principle than is given by the combination of conservation laws, with the principle of no non-physical energy, as put forward by Papineau, namely Gibb's principles of 'physical affectability' and 'redistribution' (Gibb, 2010, p.364). Following from this, I will show one way to reject sparse causal relation assumption. I will depart from Gibb somewhat, however, in arguing that one

does not *need* to reject the homogeneity[15] of mental and physical causes to doubt that physical affectability is an uncontroversial principle.

My fifth section, will return to the role of physiology in Closure, asserting that independent mental causes, of the sort ruled out by Papineau, are best understood as invisible to physical science. Further, I will argue that, even if one can devise a suitably modest Closure claim, it will necessarily fail to rule out the possibility of independent mental causes, giving the example of "in-line" mental causation. Complementing this defence of "in-line" invisible mental causation, I will also defend the possibility of "physically undetectable" simultaneous causation, against Peacocke's objection that this sort of causation is effectively overdeterministic, and thus ruled out elsewhere in the Causal Overdetermination argument. To dispute this, and to cast doubt on Papineau's Physiology argument, I will draw attention to an implicit, but seldom addressed, divergence between explanatory adequacy and causal sufficiency. This section will also make the argument for plausible mental causal efficacy even where an empirically acceptable form of Closure is granted to be true, given the aforementioned distinction between explanatory adequacy and causal sufficiency.

The final section of this paper moves to consider overdetermination. With the NUO and causal singularism dichotomy in mind, I consider standard moves in relation to attacking the overdetermination premise. I demonstrate, then, how NUO plays a pivotal role in undermining the intuitive force of the premise alongside more formal argumentation. Finally, I consider some objections to these responses, and how NUO relates to them.

[15] A homogeneity condition simply holds that mental and physical causes are the same kind of thing, namely Kimean events, as will be described in the first section.

1.1.1 Rejecting the Conclusion

In response to the Causal Overdetermination argument, the dualist could reject that irreducible mental properties have any physical effects. This amounts to a rejection of the "Mental Causation" premise. Dualists can avoid the token-identity conclusion of the argument, by accepting epiphenomenalism about mental properties. While some might be willing to bite the bullet, this is not a popular strategy. It has worrisome implications for moral responsibility, and entails the counterintuitive position that human agency is non-existent. Therefore, arguing that mental events, like our beliefs and intentions, have no effect on the world, tends to be the last resort for dualists (Crane, 1995, p.17).

Another way to avoid the conclusion is to accept systemic overdetermination. This allows one to contend that the mental cause acts in addition to the physical cause, to bring about the same physical event. Some 'belt and braces' theorists adopt a strategy of disputing both overdetermination and Closure (Mellor, 1995, p.103). For now, it will suffice to say that many take embracing overdetermination to be an ad-hoc or peculiar dialectical move (Lowe, 2000, p.572). This is not to rule it out, but for current purposes, I will not be concerned with overdetermination.

One final option is to reject homogeneity (Crane, 1995, p.17), or the principle that mental events are causes *in the same way* that physical events are. This principle seems to be implicit in the argument, but it is necessary for the Overdetermination argument to go

through. Without homogeneity, one can find a place for mental causes outside of the casually closed chain of events that the argument imagines. For Dretske, this involves taking mental causes to be 'structuring' rather than 'triggering' causes (p.17). Lowe combines the rejection of homogeneity with a rejection of Closure, by positing that mental events cause physical *facts*. Gibb makes a similar move, positing that mental causes are double-preventers (Gibb, 2015.c). I will consider this move in my final section, although I will argue that Lowe's strategy in particular can work *without* rejecting homogeneity.

1.2 Causal Closure

1.2.1 Formulating Causal Closure

Although, for sake of simplicity I presented one Closure principle within my initial formulation of the argument, there is not just one Closure principle, but a variety which have been endorsed by different philosophers. As will be argued, some are far too strong to have any empirical support, or to avoid claims of question-begging. Some are formulated in ways which avoid talking at all about causes. First, some of these principles ought to be mentioned.

Papineau: 'All physical effects are fully determined by law by prior physical occurrences' (2001, p.8).

'[It is never necessary to] leave the realm of the physical to find a fully sufficient cause for [physical] effects' (p.8).

Spurrett and Papineau: 'All physical effects are due to physical causes' (1999, p.25).

Gillett: 'All microphysical events are determined, insofar as they are determined, by prior microphysical events and the laws of physics' (Gillett, 2002, p.98).

Sturgeon: 'Every physical effect has a fully revealing, purely physical history' (Lowe, 2000, p.573)

'No physical effect has a non-physical cause' (Lowe, p.574)

Noordhof: 'Every physical effect has its chance fully determined by physical events alone' (Lowe, p.574)

Lowe (and others): 'Every physical event which has a cause has a sufficient physical cause' (Lowe, p.574)

Papineau's first formulation, and that of Noordhof and Gillett, employ the concept of *determination*, in that physical causes fully determine, or fix the chances of, some physical effect. Papineau's second formulation, and Sturgeon's first, both employ notions of *physical history,* suggesting that every physical event is explicable by reference to a purely physical past. All of the other formulations talk explicitly about sufficient causes and effects, or *events*. The idea of a complete and explanatory physical history is an important underlying assumption of Closure, as will become clear when looking at Papineau's argument from physiology. However, it seems preferable to be explicit about the causal nature of that

history, since Closure forms part of a *causal* argument for physicalism. Therefore, 'sufficient cause' formulations are preferable, if only for clarity's sake.

Determination accounts have been offered largely to avoid objections regarding low-level, or quantum-mechanical indeterminism (Lowe, 2000, p.575). Nothing that I have to say going forward, however, rests on this sort of worry, and determination accounts certainly have no greater, or substantially different, ontological or evidentiary commitments[16] than the Closure formulation I intend to focus on, so talk of determination will be set aside.

On the other hand, formulations like Sturgeon's 'No physical effect has a non-physical cause', *do* have additional commitments. This principle does not just claim that there is a complete physical history, or that there are always sufficient physical causes. It rules out any non-physical cause (Lowe, p.574). This is evidently a stronger claim than required by the Causal Overdetermination argument, since it eliminates the need for the Overdetermination premise. Further, it is unsupported by empirical evidence. There may be support for claiming that we cannot *observe* any non-physical energy, and this sort of claim will be countenanced in the discussion of Papineau shortly. Even this, however, does not ground a claim that there are never non-physical causes. The evidentiary burden is stacked against this sort of formulation, and it is unnecessary anyway, since a weaker claim, as mentioned, will suffice for the Overdetermination argument.

[16] At least, in the relevant areas of discussion. They may well be more in line with our best physical understanding.

I take it, therefore, that the most apt formulation of Closure is 'Every physical event which has a cause has a sufficient physical cause' (Lowe, p.574). It captures the essence of Papineau and Spurrett's focus on sufficient physical causes, but allows for the possibility of uncaused physical events, which are non-problematic with regard to mental causation, regardless.

1.2.2 Arguments for Closure

I will now consider three arguments which attempt to motivate this Closure principle. The first is a general strategy suggesting Closure is an obvious a-priori implication of physical science, or a necessary condition of believing in the completeness of physical science. After addressing this approach, and offering an argument against it, I will move on to presenting Papineau's empirical arguments for Closure, from fundamental forces and physiology.

1.2.2.1 The A-Priori Defence of Closure

The argument for a priori acceptance of Closure could run like this:

(1) Physics describes 'basic entities' and is concerned with 'the bottom level of description of reality' (Vicente, 2006, p.153).
(2) Since physics is basic, in the sense of dealing with the most basic properties, it must be 'explanatorily comprehensive' (p.153), or explanatorily *complete*.
(3) Physics, if complete (as it is established to be, in (2)), can provide a fully physical causal history for every physical event (Closure).

Montero offers an even more basic argument of this kind, in reverse:

(a) 'If Closure were false, then physics would be necessarily incomplete' (Montero, 2003, p.177).

(b) 'Physics is not necessarily incomplete' (p.178).

Conclusion: Closure is true.

There are various points of contention in both arguments. The leap from (1) to (2) in Vicente's argument requires some sort of additional premise to validate the move from a "basic" description to a "complete" or "comprehensive" description. If some form of emergentism is conceivable, or not clearly *false,* then one might suppose there are properties operating at higher levels than the most basic physics[17], and that physics does not explain all causally relevant properties

Both arguments, however, rely on one common assumption. This underlies the third premise of the first argument, and Montero's premise (a). This is that the truth of Closure is coextensive with the possibility of a complete physics (Montero, p.177). Completeness here means an explanatorily complete theory of 'absolutely everything' (p.178), at least of everything physical. Closure does seem to necessarily entail Completeness, since it is prima facie clear that one could provide a complete physical account of a world which is physically

[17] Emergent chemical, or biological properties, for example.

causally closed. However, there is far less motivation to suppose that the entailment runs in both directions.

Montero points out that completeness can be defined in terms amenable to all sides of the mental causation debate (p.178). If one takes it that the aim of physical sciences is to give a complete explanation for microphysical events, and for physiology (including neural and bodily movements), then this aim could be equally fulfilled by accommodating mental causes, if and when they were discovered (p.178). This is not to say that there are independent mental causes, but there seems to be no reason why a complete physics would have to rule them out, short of a question-begging presumption that only physical causes are acceptable causes, or explanations of physical effects (Bishop, 2006, p.47).

There is a further move available to the a-priori Closure proponent. It could be argued that there is independent reason to refuse to accept non-physical causes in a complete physics, in light of the way that physical causation is observed to operate, per causal singularism. This would be an 'energy flow or momentum transfer' view of causation (Kim, 2007, p.17), which holds that physical causal effects require some form of energy transfer. On this basis, non-physical causation seems to be ruled out, necessarily. However, this utterly undermines the Causal Overdetermination argument, since an essentially physicalist causal account is strong enough to not require Closure in the first place. Even aside from that, such views of causation already implicitly accept physicalism as I have suggested, allowing it to determine the metaphysical conditions for acceptable causes. Some dualists, like Hart, who accept the possibility of psychic energy as a conserved quantity, might find this to be perfectly acceptable (Lowe, 2000, p.572). However, such a view of causation clearly "stacks the deck"

in favour of physicalism and should not be readily accepted without further non-physicalist motivation.

Before moving on to Papineau's arguments, I will make two further comments regarding Closure and Completeness. Bishop notes that many physical laws set up 'typicality conditions' (Bishop, 2006, p.47), rather than giving an idealised and consistent picture of causal powers within a given domain. Typicality conditions determine what will occur under typical, or standard, circumstances (p.46). They do not address *every* instance. Therefore, one can imagine that, even a complete physical account might offer variations on a theme of "Under typical circumstances, there will be a sufficient physical cause for x physical effect". Typical circumstances might simply not include instances where minds are involved. Therefore, an explanatorily complete physics, which makes no particular mention of independent mental causes, would still not do anything to rule them out. Certainly, if one holds the Sparse View, one would be motivated to minimize commitments, especially to the potential for non-physical properties or events. However, it is unclear why one need accept the sparse view, part or whole. Even if one wishes to interpret physical laws more strictly, it is hopefully now less clear that the move from Completeness to Closure is not straightforward.

Vicente also notes that trust in the *eventual* completeness of physics might be unwarranted. Our current confidence in the prospects for Completeness are divergent from historical interpretations of physics, which were happy to admit of gaps in the causal story (Vicente, 2006, p.155). 'Epicurean and Leibnizian' physicists were sceptical of hopes for finding physical causes for each effect (p.155). Papineau argues that there is no reason to suppose

our current faith in Completeness will change (1999, p.27). Yet Montero argues that physics has surprised us in the past, and there is no clear or obvious reason to suppose that it will not again in the future (2003, p.179). I am inclined to agree, given the historical precedent of scepticism about Completeness. At very least, there is no reason to suppose it is *obvious* or *certain*, that every causal gap will be plugged with a physical event. Yet, there is this tendency to assume that physics will eventually be complete, and will give physicalism everything it needs to fill out the causal picture. This might be correct, but more argument is required to get there.

1.2.2.2 The Completeness of Physics: The Argument From Fundamental Forces

Papineau believes he has such an argument for the completeness of physics and Closure, with his two inductive arguments, from fundamental forces and physiology.

The fundamental forces argument is elegantly straightforward. All past 'special forces' have been found to reduce to a small set of basic physical forces, which follow the conservation of energy and momentum (Papineau, 2001, p.27). Vital forces, the special forces hypothesised to give life, were eventually reduced to this standard set of conservative forces, observed at work, not just in organic bodies but everywhere else (p.24). This pattern has been apparent with many so-called special forces and sorts of energy in the past. Papineau argues that it is just highly unlikely that there are any special forces, including mental ones, which do not reduce to the conservative ones which have been consistently found in physical (p.14/p.30). Furthermore, Papineau asserts that if one did posit irreducible mental forces, there would be principled reason to suppose they would follow the same

laws of conservation as the basic physical forces, which gives reason to reject any chance of their interaction with the conservative "closed-system" of the physical domain (p.30).

Elsewhere, Papineau offers the completeness of 'quantitative phenomena', like size, shape and motion, in place of fundamental forces and conserved energy (Papineau, Spurrett, 1999, p.26). A similar thought, that the domain of these basic physical entities, is causally closed, applies.

One similar objection to that offered against a priori motivations can be applied here as well. The current incomplete state of physics does not offer epistemic grounds to suppose that future physics will reduce all forces to conservative forces (Daly, 1998, p.197). Papineau even admits that future physics will likely need further explanatory categories than currently offered (Papineau, 1991, p.38). There have been some notable reductions in the history of science, like the reduction of vital forces mentioned earlier. However, the higher-level realisability of various phenomena, and the possibility of emergent properties[18] (Vicente, 2006) speaks against the hopes of universal theoretic reduction. This is not to say that reductions to conservative forces is in principle impossible, but it is not *certain*. If the supposition that future physics will not involve reference to special forces, or that it events falling under conservative physical laws will have complete physical explanations, is not supported by more than a mere belief that future physics will also feature purely conservative *physical* forces[19], then it begins to look more like a prejudice than a well-founded principle. We can return here to consideration of the Sparse View. There seems no

[18] This is discussed in greater detail in 3.9.3.

[19] Even if true, I have already noted that certain philosophers, like Hart (1988), are happy to posit fundamental mental, or psychic, forces as conservative.

empirical reason to make the inference Papineau makes here, yet with a prior assumption that the only relevant causal relata will be physical and interact according to physical laws without mental intervention, it becomes intuitively easer to put credence in such a prospect. We need not.

1.2.2.3 The Argument from Physiology

Papineau, however, does not stop with fundamental forces. He offers, in my view, a stronger complementary argument which considers evidence from physiology. He states that the enormous amount of 'neurophysiological research' carried out with regard to bodily and brain processes has revealed no added energy, or mental forces at work (Papineau, 2001, p.31). Certainly, there are functionally complete explanations for many of our bodily processes, including those which seem to involve human agency (reaction, movement and so forth) which feature entirely adequate physical explanations (p.31). All of these organic processes are 'fully accounted for' (p.27). This is not to say there *cannot* be mental causes, but there is no evidence for them, nor for any sort of non-physical energy in humans.

There is, therefore, no explanatory role for independent mental forces, or events, within the causal story. I will return to the Physiology argument in my final section, to point out an underlying flaw in this reasoning which relates to this focus on 'explanatory adequacy' (Crane, 1995, p.6). Beforehand, however, I will consider an objection to Papineau's two arguments which disputes the extent to which they independently support Closure.

1.2.3 Conservation Laws and Closure

Papineau's arguments, if successful (and I will grant this for now), can be taken to ground two principles, namely *Conservation*, and *Energy*. Conservation is supported by Papineau's argument from fundamental forces.

Conservation: 'Every physical system is conservative or is part of a larger system that is conservative (where a system is conservative if its total amount of energy and linear momentum can be redistributed, but not altered in amount, by changes that happen within it)' (Gibb, 2010, p.367).

The Energy principle draws on the argument from physiology, which shows that any non-physical energy is, at best, unlikely.

Energy: There is no non-physical energy (p.364).

Papineau takes it that Closure follows from these principles. *Conservation* would be violated by mental interaction, unless there is non-physical, and *conservation-governed*, energy. Yet, from *Energy*, mental energy is ruled out. Gibb argues that Closure does not follow from these two principles. Closure is, of course, the principle that every event which has a cause, has a sufficient physical cause.

It is noted by Gibb that, insofar as *Conservation* and *Energy* can be taken to entail Closure, they entail the strongest Closure principle, that there can be no non-physical causes of physical events (Gibb, 2010, p.369). Since nothing rests on this point, and that the strongest form of Closure entails the weaker version of Closure that Papineau wants anyway, this will not be a point of contention.

1.2.3.1 Physical Affectability and Redistribution

Gibb's argument aims to demonstrate that, to get Closure, Papineau requires at least two additional principles to be true, namely *Physical Affectability*, and *Redistribution* (Gibb, 2010, p.370). I will not delve into Redistribution in this section, since it is simply the claim that 'Redistribution of energy and momentum cannot be brought about without supplying energy or momentum'[20] (p.374). I think Gibb is right to suggest that this is a necessary condition for Closure, but I do not view it as problematic, in the same way that *Physical Affectability* is. First, I will explain why Papineau's premises are insufficient for Closure.

The central problem, for Gibb, is that neither of Papineau's premises feature any sort of explicit causal claim, but they do feature some causal assumptions (p.370). The causal relation assumption restricts the ways in which a property, event, or other relata, could be a cause, to certain sorts of physical interaction. Gibb terms this Physical Affectability instead of Causal Relation, but the relation between the assumptions Gibb and I identify is apparent.

[20] This *could* be problematic for Closure if accounts of independent mental events as "prompters" (p.374) which Gibb considers, turn out to be coherent. This will be addressed in a later chapter.

To get from *Conservation* to Closure, the manner in which causes can affect physical systems must be limited. Conservation laws prohibit the addition or creation of energy in closed physical systems, unless those systems are sub-parts of larger conservative systems (p.370). However, it does not rule out the possibility that 'energy and linear momentum can be redistributed' (p369/370) without any change in the amount of energy within the system.

It is necessary to include an additional principle to derive Closure, since *Conservation* and *Energy* do not block the possibility of causal effects upon physical systems which do not involve the change or redistribution of energy or momentum (p.370). Therefore, Gibb proposes this:

Physical Affectability: 'The only way that something non-physical could affect a physical system is by (1) affecting the amount of energy or momentum within it, or (2) redistributing the energy and momentum within it' (Gibb, p.370).

This amounts to the claim that the only way for an event to have *causal impact* is for it to affect or redistribute energy or momentum. One might grant this is a necessary underlying assumption of *Conservation*, or even a necessary additional principle, but take it to be so totally uncontroversial as to cause no problems for the move from *Conservation* to Closure. Indeed, energy and momentum changes are ubiquitous in causal chains of physical events. It is hard to conceive of causation which is not constituted by these sorts of change. To the

extent that *Physical Affectability* is necessary for, or even exhaustive of, event-causation, the need for this principle is not problematic for Papineau.

However, Gibb suggests one reason for scepticism about *Physical Affectability*. She points to the plausibility of dualist accounts which offer ways for mental events to affect the physical, without affecting the amount, or distribution, of energy or momentum (Gibb, 2010, p.372). In particular, she cites Lowe's account of mental causes, which positions mental events as having the effect of making certain physical effects non-coincidental (p.371). The essence of this position is that the convergence of disparate physical events, neural and bodily, on some particular bodily action, like moving one's hand, can appear coincidental with regard to the physical causal story alone (p.371). Mental events, intentions in particular (Lowe, 2000, p.579), fill the causal role of rendering this convergence non-coincidental (p.579).

There are differing ways of interpreting Lowe's position, regarding whether he takes mental events to just occupy a unique causal place, linking different causal chains which end up converging on a given action, or whether he takes mental events to purely cause certain physical *facts* (p.582). The latter assumes a novel ontological difference between different forms of causation, namely event-causation and fact-causation. The physicalist is not obliged to accept this distinction. Equally, one might be sceptical with regard to treating coincidence, or the capacity to render an event non-coincidental, is anything like a causal notion. It seems, *prima-facie* an epistemic or explanatory feature of certain mental events. If we understand Lowe's claim to be shifting the nature of the relata, then we have an instance of implicitly rejecting the causal relata assumption as a means to undercut the argument. If there are different sorts of relata at work, then closure is not violated, since

the physical causes are sufficient for the physical effects, and overdetermination does not seem to occur, since the facts and events are distinct, and determined separately. Simply rejecting this assumption offers a way to reject both premises at once.

I do not wish to say that Lowe's account is necessarily wrong. Indeed, the mere plausibility of his account might suffice to cast doubt on *Physical Affectability*. To this end, it is at reasonable to hold up an account like Lowe's as sufficient for denying the obvious nature of *Physical Affectability*. However, I think one can make a case against Physical Affectability which does not carry the ontological and argumentative burden involved in invoking Lowe's account.

I have two sorts of strategy in mind. The first focusses on certain plausible interpretations of quantum theory. The second addresses the notion of background conditions for causation.

There is evidence for true 'randomness' (Grygianiec, 2016, p.12) or non-determined variability in the state and position of certain subatomic particles, within modern quantum theory. However, this sort of randomness can be probabilistically weighted, such that the relevant chances of a certain subatomic state of affairs at time t2 are determined at time t1[21]. Certain plausible interpretations of quantum mechanics even position conscious observation as a factor in fixing the probability of certain quantum events, like the collapse of the wave-function (Montero, 2003, p.179). However, the point I wish to make clear here does not require any commitment to, or claims about, the causal powers of minds with

[21] Not deterministically, however. It is the weighting of the relevant probabilities that may be determined, not the outcome itself.

regard to microphysical or quantum phenomena. It does not even require that this sort of quantum theory turn out to be *true*. All I am seeking to establish is that there are, in fact, physical theories which consider sorts of genuine physical change that do not involve redistribution of energy or momentum. It might involve a displacement of certain particles, but even this is not necessary. One might suppose that there has been a real change in the world just if there has been a change in the probability of some later event occurring.

To illustrate this, we can imagine two possible worlds, X and Y. At X, there is an 80% chance of some particular quantum change at time t2[22]. At Y, no other features of the world are different, but there is a 60% chance of the same change at time t2. It is, I think, reasonable to assert that these worlds are different in a physical respect, or with regards to certain physical properties even if only their dispositional properties.

In addition, consider the following from Sider:

'An effect can depend counterfactually on both the instantiation of mental and of physical properties, can be subsumed under both purely physical laws and psychophysical laws, can have its probability raised by the instantiation both of mental properties and of physical properties, and can be related by a primitive causal relation to both a mental and a physical cause' (Sider, 2003, p.721-2).

[22] I leave aside, here, whether worlds differing in this way, holding all else equal, are nomologically possible. This serves only to illustrate the principle regarding probability.

This is, naturally, a major oversimplification of the sort of physical theory I am referencing, but the principle is hopefully clear. Physical Affectability states that changes to physical systems must involve alteration of energy or momentum. However, I have argued that physical change can involve just the alteration of possibilities or chances of later events in an indeterministic world. One could even move away from quantum mechanics entirely and simply suppose that indeterminism at *any* level opens the door to differential possibilities, and the potential for mental events to weight those possibilities. A commitment to complete physical determinism would rule out this sort of response, but it is not clear why someone, especially a dualist sceptical about Closure, should accept determinism *anyway*. I am not committed to the view that dualist mental interaction might involve alteration of possibilities, at the level of quantum mechanics or at the macro-level, but it is one strategy that is available to the dualist (Meixner, 2008, for one such treatment of mental causation), and which undermines *Physical Affectability*.

The second strategy for undermining *Physical Affectability* highlights the possibility of mental causes as background conditions for certain sorts of bodily effect (Gibb, 2015.a, p.631).

In this sense, mental causes have their effects not in virtue of altering or redistributing energy, but through providing the necessary conditions for the occurrence of some physical event. This draws a distinction between the cause, as an event which causes another, and 'those events which are mere background conditions necessary for the relevant causal relation to take place' (p.631). An example might be lighting of a match, where the match is the cause, but the presence of oxygen and dryness of the match are background conditions,

necessary for the lighting (p.631). Gibb herself dismisses this sort of approach, on the grounds that causation is an objective relation between certain prior events, and an end-result, whereas the division between causes and background conditions is artificial, or subjective (p.631). We view certain events as causes, and certain events or properties as providing background conditions, just because of our perspectives, which lead us to see certain events as active, and others as passive. However, in reality, the passive presence of the oxygen was just as much a part of the cause as the striking of the match. There might be different "causal chains" or seemingly distinct events which lead to a given effect, but these can be understood as joint causes. They should not be arbitrarily divided into causes and conditions.

However, I think this dismissal is premature. The match analogy is not generalisable to all sorts of causal affect. When talking about the various physical properties and events which have causal input, it makes sense to view these inputs in the same way, as contributory causes (p.631), at least without good reason for thinking otherwise. All of the physical antecedents of the ignition are justifiably thought of as the same kind of cause. However, taking property, or substance, dualism seriously gives us at least one reason *to* think otherwise[23]. If mental events are distinct, then they instantiate different sorts of property[24], that are non-physical in nature, and one might therefore suppose that these sorts of property instantiations alter bodily events in non-physical ways. Certain sorts of bodily action are *intentional*, from picking up a cup, to moving in a given direction. It seems relatively uncontroversial to say that an intentional action is fundamentally distinct from a

[23] I assume, *at this point*, that these positions do deserve to be taken seriously, since the argument from Causal Overdetermination is aimed at showing such positions to be false.
[24] This would be a minimal claim of most conventional construals of both property and substance dualism.

non-intentional action. Prior mental events are necessary for intentional action. I do not deny that the reductive physicalist can offer a coherent story here, which identifies intentional mental events with physical events. However, the contrary possibility, that there are mental events which imbue intentionality, without affecting the distribution of energy or momentum, is not obviously incoherent either. At very least, one might suppose that certain mental events are necessary conditions for certain sorts of physical event, like intentional action, without being direct causes of that action. This line of argument will be developed further in the final section, which considers causal relevance in contrast to sufficiency.

In this section, I have advanced two objections to the *Physical Affectability* premise, casting doubt on its supposedly obvious nature, without relying on the ontologically cumbersome notion of "fact-causation". My prior critique, in the third section, of Papineau's argument from fundamental forces, also gives reason to view *Conservation* as suspect. Having established that Closure is founded on certain undermotivated principles, I will now move on to consider ways that independent mental events might be causally relevant, even when Closure is accepted. Connected to this critique, I will also offer reason to think that the *Energy* principle, might be as undermotivated as *Physical Affectability* and *Conservation*.

1.2.4 Explanatory Adequacy and Causal Sufficiency

I am now in a position to return to Papineau's argument from Physiology, taken to establish *Energy* specifically, and Closure more broadly. It establishes Energy by ruling out mental energy from explanations of physiological phenomena. Since no mental energy has been

observed at work in bodies, and since “special” mental events have been unnecessary for offering explanations of bodily behaviour, Papineau takes it that we should dismiss the possibility of independent mental causes, or energy (Papineau, 2001, p.31). Crane notes that this is an appeal to the ‘explanatory adequacy’ (Crane, 1995, p.6) of physics, or physiology in this case. These empirical explanations, positing certain physical causes, are adequate for the task of explaining physical effects, in every bodily instance.

I will now suggest, and defend, two ways in which mental causation is invisible on even an adequate physical analysis. These will be cases of in-line causation, and simultaneous causation.

1.2.4.1 In-Line Causation and Simultaneous Causation

Closure states that every physical event has a sufficient physical cause. This claim is entirely consistent with two different, mutually exclusive causal paths, (1) and (2).

(1) P1-> P2

(2) P1->M->P2

In both cases, P2 has a sufficient physical cause (Lowe, 2000, p.582). However, in (2), P2 *also* has a sufficient mental cause (p.582). Positing a causal chain like (2) for any bodily event would not be a weird or ad-hoc move by the interactionist dualist. Any such dualist will already be inclined to support the close causal connection of brain events (as P1 might be), mental events, and bodily actions.

Furthermore, the possibility of (2) should not undermine the explanatory adequacy of physiology, even though it does pose a problem for Closure. It remains the case that there is an explanatorily adequate explanation for how P2 came about. The omission of M is not problematic for a pragmatic scientific account, since P1 does all the necessary explanatory work. However, this demonstrates that Closure, even when it is nominally correct, does not necessarily succeed in eliminating causally sufficient mental causes. A lack of evidence for M should not be taken to mean that M does not exist (Montero, 2003, p.185), and a lack of evidence for M should not imply *Energy*, or Closure.

One way to avoid the possibility of (2) would be to embrace a stronger form of Closure, like the claim that every physical effect has an *immediate* physical cause (Lowe, 2000, p.579). However, this is far too strong to find substantive empirical support. As Lowe puts it, temporal ordering is dense (p.579/580), and to suppose *immediate* physical causes commits one to a claim about the contents of every fine-grained time-slice. This is clearly worrisome, since there is no feasible empirical evidence for physiological closure (Bonjour, 2010, p.6).

However, even if one is willing to bite the bullet and embrace an empirically unsupported "immediacy" Closure thesis, there is another way in which mental causes might be invisible to physiological explanations. I have in mind cases of simultaneous causation. P1 might cause M, and both P1 and M might be jointly responsible for causing P2 (Lowe, p.577). The physical causal story might point to P1 as the cause of P2, whilst remaining silent on M. One could simply rule out simultaneous causation as absurd, impossible, or unlikely, but any statement like this would constitute another 'substantive claim' (p.577).

Simultaneous causation may save mental causation and present a Closure problem for the Causal Overdetermination argument, permitting that mental events are empirically undetectable yet proximate causes of physical effects. However, Peacocke argues that, while simultaneous causation is not incoherent or problematic with regards to Closure, it is nevertheless ruled out by the prohibition against causal overdetermination itself (Peacocke, 1977, p.136). Putting aside for now that overdetermination may be less problematic than often supposed[25], I will dispute Peacocke's claim that physical-mental simultaneous causation is a sort of overdetermination. Apart from defending the possibility of simultaneous causation, my response to Peacocke will consider how the Sparse View is involved in such arguments. Peacocke's identification of simultaneous causation with overdetermination hinges on the causal sufficiency of the physical causes in question (p.134). Let us imagine that there is an explanatorily adequate link between P1 and P2. For Peacocke, *if* there is an explanatorily adequate explanation for P2 offered by neurophysiology in terms of P1, this must mean that P1 is causally sufficient for P2 (p.134). Since P1 is causally sufficient in virtue of being explanatorily adequate, M is an overdetermining cause and thus ruled out by the Overdetermination premise.

There is an assumption here that is also implicit in Papineau's arguments, and which we need not endorse. This is the assumption that explanatory adequacy entails causal sufficiency. To put it another way, where X is explanatorily adequate for Y, this means that X is causally sufficient for Y. This rules out the possibility of empirically invisible mental causes

[25] This is discussed in 1.3.

(Lowe, 2000, Robb, 2018) which are explanatorily irrelevant to physics. When I claim that this assumption is present in Papineau's work as well, I mean to take issue with the move from the explanatory adequacy of physics (Crane, 1995, p.6) to the causal sufficiency of physical causes for physical effects. If it can be argued that explanatory adequacy does not entail causal sufficiency, then Closure should not be thought to follow from the success of physics or physiology, contra Papineau. Regarding Peacocke's specific case, there would be no reason to suppose that simultaneous causation need be ruled out.

Before I offer any objection, I will make one comment on motivations for this assumption. Certainly, it is intuitive to think that explanatory adequacy entails causal sufficiency. In many cases, it seems obviously true. If "Jack dropped the ball" is explanatorily adequate "The ball hitting the floor" (in concert with the laws of physics and so on), then it is reasonable to think that Jacks' dropping of the ball is causally sufficient for the ball hitting the floor. In chemistry, if the creation of carbon dioxide and water in a given reaction is explained adequately by the burning of some methane in an oxygen-rich environment, then that reaction was obviously causally sufficient for that creation.

If there is reason to suspect that the entailment from adequacy to sufficiency is not necessary, this will be a problem for Closure. However, there is no reason to suppose this entailment is necessary, or universal. The relation between explanatory adequacy and causal sufficiency is not an uncontentious one in the philosophy of science (Boyd, 1991, p.755/756). A "wave of selling in the Stock Market in 1929" was sufficient to cause the Great Depression (p.756) but it should not be supposed that the "Selling" event is an adequate explanation for the Depression. However, this would only establish that causal

sufficiency and explanatory adequacy are not co-extensive, as this is a case where we have causal sufficiency without explanatory adequacy. It does *not* establish that explanatory adequacy does not entail causal sufficiency, which I shall now attempt to do.

I will offer three examples of explanatory adequate explanations which do not have sufficient causes as their object. First, we can recall the in-line mental causation model offered earlier. With (2) P1->M->P2, M can be viewed as a sufficient cause for P2. And P1 is a sufficient cause for M. However, if one is looking for an adequate account of how the token event P2 came about, it would be physically, and explanatorily sufficient to explain the occurrence of P2 by reference to P1. Of course, it may be objected that this stipulates just the sort of case which is in contention. It may, perhaps, be helpful, to offer a more common-sense example.

"Snow is falling" is explanatorily adequate, at least in quotidian contexts, for the later state of affairs "There is snow on the ground". However, it is not causally sufficient; since other factors, from the constancy of gravity at ground level, to cold, dry ground, to an absence of snow-ploughs are part of the sufficient cause, or causal conditions, for the "Snow on ground" state of affairs coming about.

It should be noted that, amongst this list, there is an absence, or *omission*, in the form of the absence of snow-ploughs. I will remain neutral on the question of whether omissions are causes. However, if that turns out to be the case, then this only strengthens my point. For if omissions are causes, then they would form part of the causally sufficient conditions in a whole variety of cases. Certainly, it is not *inconceivable* that omissions might be part of the set of causally sufficient conditions for an event. However, omissions (most of the time) will not be part of a legitimate, explanatorily adequate story for the occurrence of any given

physical event. In (2), for instance, an explanatorily adequate account would *definitely* not include the absence of a preventing-effect which would stop P2 from taking place.

Therefore, it seems like explanatorily adequate stories need not contain all parts of the sufficient cause for a given event. This, I think, gives good reason to suppose that simultaneous causation need not be treated as a species of overdetermination. This also undermines the common move in the literature from explanatory adequacy of physics generally, to Closure. The clearest example of this regards Papineau's argument from physiology. When the Physiology argument asserts that the success of physiological explanation entails the causal sufficiency of certain bodily, physical causes, this entailment is not as certain as might be supposed.

1.2.5 Final Thoughts on Causal Closure

The closure premise is susceptible to various sorts of challenge, whether by modifying the causal relata as Lowe suggests, or through broadening the ways in which a cause might cause, as I and Gibb concur, through a rejection of the causal relation assumption.

This paper began with an outline of the Causal Overdetermination argument, which laid out its premises, including Closure. After this, I briefly addressed certain metaphysical commitments of the argument. I highlighted various options open to the dialectical opponents of the argument, including the rejection of Overdetermination, and Mental Causation.

In my second section, distinct variations of causal closure principles were examined. I offered motivations for putting aside probabilistic, "physical history", and overly strong variants of Closure, opting to focus on the following: "Every physical event which has a cause has a sufficient physical cause".

In the third section, I considered both a-priori and empirical defences of Closure. A-priori arguments held that Closure was entailed by the *in-principle* completeness of physics. I disputed the link between completeness of physics, and Closure. I also offered some initial reasons to suppose that Papineau's motivations for Closure were themselves undermotivated. In both cases, I highlighted certain incorrect historical and predictive assumptions regarding the truth of Closure.

Section four introduced Gibb's argument against Papineau's move from *Conservation* and *Energy* to Closure. Incorporating her insight that these principles fail to entail Closure, I offered arguments for the under-motivation of *Physical Affectability*, which did not rely on the ontologically weighty rejection of mental-physical homogeneity *pace* Gibb.

My fifth section returned to the consideration of Papineau's Physiology argument, as well as the more general form of explanatory adequacy argumentation for Closure. I argued that even a reasonable Closure principle, derived from the success of physiology, cannot rule out sufficient mental causes, or the possibility of joint mental and physical sufficient causes. I also defended the possibility of simultaneous causation against the charge that is was necessarily a species of overdetermination. To do this, and to cast further doubt on the Physiology argument, I highlighted the potential for certain events to be explanatorily

adequate, without being causally sufficient, for certain effects. This allowed me to motivate the possibility of genuine simultaneous causation and undermine Papineau's leap from the explanatory success of physiology to the sufficiency of physical causes.

In general, this chapter has addressed various problems with the Closure principle, and has at times pointed to ways that the underlying sparse assumptions might be rejected by arguments against the view. I have critiqued Closure on several levels, including addressing the flaws in its underlying motivations, and highlighting additional assumptions besides Conservation and Energy which are required of Closure. I see no knock-down argument against Closure. However I would argue that the sparse assumptions are required to support Closure. Further, the doubt cast on the connection between Closure, explanatory adequacy, and Completeness, give substantial reason to doubt the principle. I have also offered reasons to suspect that no reasonable formulation of Closure can eliminate the possibility of sufficient mental causes *anyway*, in light of in-line causation, mentioned in section five. This could mean that Closure simply does not fill the role which the Causal Overdetermination argument requires of it, regardless of the truth of the principle.

With closure better understood, and some work done to establish the sparse assumptions as they relate to the argument, we move on to consider overdetermination, which I take to be far more substantially grounded, as a purely metaphysical principle, in intuitions. These intuitions, insofar as they rely on very plausibly erroneous sparse assumptions, can be rejected.

1.3 Systemic Overdetermination: Causal Singularism and Coincidence

1.3.1 Overdetermination, The Sparse View, and Funkhouser's Taxonomy

Overdetermination, as a reminder, is any case in which one event has two or more causes, where any one individual cause would have been sufficient to bring about that exact event (Schaffer, 2003, p.23-24, Bernstein, 2016).

As discussed in the first section, I take it that a defense of NUO should go some way to undermining prima-facie dismissals of systematic overdetermination. However, even with a starting-point assumption which does not render the suggestion of overdetermination "weird" or "ad-hoc", there is still a need to account for how overdetermination might arise in token cases, and to explain differences between the intuitively peculiar examples of overdetermination so often provided, and what seems to be occurring in the case of overdetermining mental and physical causes.

In what follows, as in my prior argument regarding Closure, I intend to demonstrate how the Sparse View, and causal singularism specifically, feature in arguments for the overdetermination premise.

I begin with a brief discussion of overdetermination itself, elaborating on some considerations not canvassed in the introduction. Having set the stage, I will canvas typical pre-argumentative responses from philosophers to the prospect of endorsing any sort of widespread overdetermination. Alongside this, I will illustrate how these responses, which I

term “reflexive peculiarity” responses, are rooted in causal singularism. Then, I return to one of the standard examples of overdetermination in the literature; the firing squad. Getting clear on the moving parts of this example will be worthwhile as I continue to outline Funkhouser’s taxonomy of overdetermination types. His taxonomy provides us with the tools to highlight a further distinction between coordinated and non-coordinated overdetermination, which I also elaborate.

With causal singularism, NUO, and an analysis of types of overdetermination in place, the case for systemic overdetermination as a response to the overdetermination argument should become clear. NUO paves the way for such overdetermination, showing it to be not inherently implausible. Insights from Funkhouser allow us to more formally explain how such overdetermination might occur, while permitting us to distinguish the truly peculiar cases of overdetermination from those which, per NUO, should not even be regarded as *prima-facie* peculiar. The mental-physical case we are focused on, I suggest, falls into the latter category. Some hopefully illuminating examples of these contrasting sorts of overdetermination are provided before I move on to consider and resolve four objections to commonplace or systemic overdetermination. Two, the “worrying ubiquity” and “coincidence” objections, are readily dealt with by appeal to NUO and the Funkhouser taxonomy, while the other two will require some additional work. With the overall picture fleshed out, and objections dealt with, I conclude that those wishing to reject the overdetermination argument have well-founded reasons to accept systemic overdetermination.

1.3.2 Overdetermination and Related Phenomena

Definitions of overdetermination and systemic overdetermination have already been provided. Overdetermination makes some causes redundant, such that not all of the truly distinct causes of some event were necessary for that event (Schaffer, 2003, p.23). However, it will be worth noting a similar causal phenomenon which should not be mistaken for overdetermination. I have in mind here multiple causation, where no apparently distinct cause is redundant and they were jointly necessary to bring about the effect. Were only some of these partial causes were present, the token event *would not* have occurred (Funkhouser, 2002, p.335). This sort of scenario, where individual causes are singularly insufficient to generate the resultant effect, is also referred to as joint causation within the literature (Bernstein, 2016, p.18).

Finally, there are pre-emption cases. In a case of pre-emption, we observe two apparent causes of some event, but one cause temporally precedes the other, ostensibly entailing that the latter is no cause at all (Funkhouser, 2002, p.338). Pre-emption is sometimes presented as an objection to the possibility of overdetermination. The argument is that supposed cases of overdetermination involve putative causes which do not occur at *precisely* the same moment, and thus the temporally prior cause looks like it pre-empts the latter. If our best examples of everyday, apparent, overdetermination do not feature causes which occur at precisely the same time, then we really have lots of cases of causal pre-emption, and overdetermination seems even more rare and peculiar. I regard Funkhouser's response to this as persuasive:

'First, there could be cases in which the causes arrive simultaneously. Second, even if they don't arrive simultaneously, the latter cause could still arrive in time to do its causal work in the same way it would have had the other cause not been present' (Ibid).

If one is concerned that pre-emption cases are distinctly non-overdetermined, then take it as a point of stipulation that future examples I provide will feature causes which are fine-grainedly temporally co-extensive. This is in line with Schaffer, who distinguishes the temporally 'asymmetric redundancy' (2003, p.23) of pre-emption from the 'symettric redundancy' (ibid) of true overdetermination cases

As noted in my introduction, it is often the case that the possibility of overdetermination ubiquitously, systemically or indeed, more frequently than as a rare "freak" occurrence, is regarded as deeply counter-intuitive. Here, I provide a brief and by no-means exhaustive list of responses to systemic overdetermination, as regards intuitive distastefulness:

"Ad-Hoc" (Lowe, 2000)

Must be "odd" (Sturgeon, 1998)

An "intolerable coincidence" (Melnyk, 2003)

Ontologically excessive (Sider, 2003)

Would be "ugly" (Merricks, 2001)

For brevity, I refer to this sort of sentiment as the “reflexive peculiarity” response. Although many philosophers will offer formal arguments against overdetermination alongside reflexive peculiarity, I think the intuition itself warrants a response before arguments are addressed.

Causal singularism, as explained in my introduction, is a starting-point assumption that causation occurs at one level of reality. Alongside commitment to some sort of Closure principle, it produces a picture of causation as occurring principally, and perhaps exclusively, at the level of either fundamental physics or thermodynamics and “macrophysics”. To hold NUO, by contrast, is to start from the assumption that causation occurs not only at that level, but at many higher levels as delineated by chemistry, biology, and certain of the special sciences. I contended that NUO better fits our linguistic practices around the ascription of causation, and our intuitions regarding cause and effect. On NUO as a starting-point, effects will frequently seem to have various notional causes, which it seems natural to further take as causes *simpliciter,* absent reasons to rule such out.

Whether these notional causes are truly causes is not settled, of course, and one may provide argument against the possibility of higher-level causation. Insofar as NUO offers a compelling intuitive picture of overdetermination, we have good reason to doubt, on reflection, that overdetermination is peculiar at all. Causal singularism, then, is what makes systemic overdetermination reflexively peculiar. The philosopher who wishes to reject the conclusion of the overdetermination argument is best served to abandon this assumption and endorse NUO.

I think NUO, and the rejection of causal singularism, paves the way for those endorsing systemic overdetermination . However, denying the apparent peculiarity of systemic overdetermination does not suffice to clarify how and when systemic overdetermination actually occurs, nor does it resolve, in isolation, intuitive problems associated with certain canonical examples of overdetermination. I provide such examples now, after which I shall show that we can distinguish worrisome and non-worrisome cases of overdetermination and, moreover, that the mental-physical case is firmly in the non-worrisome category.

Let us consider four examples in a little detail, before examining what separates them.

1: The Firing Squad

A man is sentenced to death and placed in front of a firing squad composed of five soldiers. At the signal of their commanding officer, all five shoot the convicted man, and he dies. Being expert marksmen, the shot from any one of the five soldiers would have been sufficient to cause the man's death (Schaffer, 2003, p.39).

2: Death By Lightening, and Gun

A woman is pursued by an assassin during a lightening storm. At just the moment that the assassin fires an accurate shot hitting the woman's heart, she is struck by a lightening bolt sufficient to kill her. The lightening and the shot were each sufficient for her death.

3: The Broken Window Pane

Two children, playing around and acting mischievously, throw individual rocks at a window, at the same time, when either rock would have been sufficient to shatter the window. The window shatters (Bernstein, 2016, p.18-19).

Schaffer (2003, p.27) and Bernstein (2016, p.23), among others, note that Example 3 may fail to count as an instance of overdetermination at all, where we individuate the events in a more fine-grained manner. The second rock may not merely cause the glass to shatter, but to shatter with more force, rendering this a case of joint causation (ibid). I provide the example only in hopes of offering a quotidian illustration of a certain pattern of cause and effect, so if one is concerned about the example as presented, instead imagine some alternative scenario in which two distinct causes bring about the same effect, *however* fine-grained that effect may be individuated. What is important for present purposes is the conceptual structure of the case.

4: The Stock Market Crash

A set of influential investors panic one day and decide to sell stocks *en masse*. The market crashes 500 points in one day (Funkhouser, 2002, p.340). The actions of investors and others cause this crash, but it is also caused by their lower-level bodily movements (ibid).

Funkhouser provides a tripartite taxonomy of causal overdetermination. My focus will be on his first two types of overdetermination. The first is 'independent causal overdetermination'

(Funkhouser, 2002, p.337). In such cases, ‘causally relevant properties’ (ibid) of distinct objects, with independent prior causal pathways, converge on the effect. Example 2 exemplifies this sort of overdetermination. The lightening and the assassins gun, and their properties which are causally relevant to the woman’s death, are distinct. Further, the gun and lightening share no proximate common cause.

Funkhouser identifies a case like Example 1 as a further instance of this sort of causal overdetermination, in light of the bullets following distinct trajectories from distinct objects (Funkhouser, 2002, p.337-8). I put this point aside momentarily, although more must be said regarding this example.

His second type of overdetermination is ‘incorporating causal overdetermination’ (Funkhouser, 2002, p.340). Here, distinct properties of the same object or entity, all serve as individually sufficient causes of the same effect (ibid). Example 4 illustrates this type; the causal higher-level and lower-level properties are distinct, but instantiated within the same objects, the investors and others. He notes that non-reductive physicalists might embrace incorporating causal overdetermination in their model of mental causation:

> ‘A relationship like that captured by the terms ‘supervenience,’ ‘realization,’ or ‘constitution,’ holds between incorporating causes’ (ibid).

His third and final type of overdetermination is ‘iterative causal overdetermination’ (Funkhouser, 2002, p.346). Iterative overdetermination 'typically involves either second-order or disjunctive/conjunctive properties causally overdetermining alongside their first-

order companions or disjuncts/conjuncts' (ibid, p.348). Specifically, Funkhouser has second-order dispositional properties in mind (ibid). I note iterative overdetermination for sake of completeness, but it will not be relevant to the following argument and I shall therefore set it aside without further elaboration. With Funkhouser's taxonomy in place, I will now introduce one further division between types of overdetermination; coordinated and non-coordinated.

Coordinated overdetermination: An instance of overdetermination is *coordinated* overdetermination, just in case there is a proximate shared past between causes which explains their convergence on the overdetermined effect.

A shared past or shared history here would mean that the overdetermining causes in question either share a cause themselves, or that one has also caused the other. This does not extend indefinitely into the past. When I specify that the shared causal history is proximate, this means that only recently shared causal history, between currently existent objects or instantiated properties. This also rules out potentially worrisome multiplication of coordinated overdetermination instances, since we can disqualify shared causes such as the Big Bang. This is because the events, objects, and properties implicated in current causal interactions did not exist at the time, nor were they proximately caused to exist by the Big Bang, either.

Non-coordinated overdetermination instances are simply those which lack this "shared causal history" feature. Example 2 is an instance of non-coordinated overdetermination. As with Funkhouser's independent overdetermination, the lightening strike and the fired gun

manifest wholly distinct causally relevant properties, within wholly distinct objects. Further, the relevant events have no causal link or shared causal progenitor. My distinction draws particular attention to the latter point. The lightening and the bullet have no shared causal history until they coincide to cause the same event. The probability of such an occurrence is extraordinarily low. It does appear, to cite Melnyk once more, an 'intolerable coincidence' (2003).

By contrast, we may consider Example 1. Funkhouser regards this as independent overdetermination, as he would Example 2. I do not object to this classification within his own taxonomy, but I take his approach to elide an important distinction here, between instances like Examples' 1 and 2. Example 1 is an instance of *coordinated* overdetermination. There is no coincidence at play; all the soldiers fired in virtue of a shared cause, an order in the form of a signal. Not all shared prior causes lead to events which coordinate upon some future effect, of course. Instances which *do* may involve reasons, as feature within Example 1, or processes which have evolved to coordinate, as within certain biological systems. The salient point is merely that we can identify such instances by reference to some prior shared cause

I posit another reason a case of overdetermination may be coordinated. The obvious, a shared prior cause, has already been noted. Besides this, we may look to Example 4 and consider cases of supervenience (Funkhouser, 2002, p.340), or grounding. I note grounding given its status as an explanatory and *building* relation (Goff, 2017, p.43). The investors are moving around and selling stock *in virtue of* relevant microphysical properties. The microphysical facts explain, at least partially, the macrophysical and economic facts

regarding the movements of the investors. Nevertheless, whether supervenience, grounding, or some other dependence relation, an intimate metaphysical relation does exist between the causes of the stock market crash. Indeed these events, the causes, cannot occur without one another (Funkhouser, 2002, p.340). Here too, then, we can state that this overdetermination is not coincidental, but coordinated.

Examples 2 and 3 seem non-coordinated. 3 features agents, but they have no shared reason to throw rocks at that precise moment, in contrast to the marksmen in Example 1, ordered to do so. In Example 3, they merely happen to fix upon the same behavior at the same time, resulting in overdetermined glass breakage. Likewise, there is no prior cause or dependence relation between the lightening and the assassin in Example 2. My coordinated overdetermination thus borrows from Funkhouser's conception of shared causal history or a shared prior cause, but encompasses cases which he would recognize as independent, and incorporating. I take all instances of incorporating overdetermination to be coordinated, given that they *necessarily* involve dependence relations of some sort:

'Independent overdeterminers can "come apart" – that is, either one of those causes could occur without the other, [but] if incorporating causation occurs at all, it is necessarily systematic' (Funkhouser, 2002, p.340-1)

Parenthetically, this quote from Funkhouser also serves to highlight a distinction in purpose between his taxonomy and my division. I am concerned with explaining the difference between cases of overdetermination which evoke negative intuitions, and those cases which seem natural and explicable. Given this, my interest is not in whether some causes

could, modally speaking, “come apart”, but only whether actual token cases share a relevant coordinating feature which non-coincidentally explains their overdetermination. Non-coordinated overdetermination involves causes lacking both features and hence any explanation for their overdetermination of the same effect.

The case of mental and physical overdetermination is, at least for the non-reductive physicalist, an instance of incorporating causal overdetermination on Funkhouser’s taxonomy. Necessarily, as consequence, it is a case of coordinated overdetermination by my definition. Further, it does not arise from any sort of bizarre contrivance of circumstances, given that, on typical non-reductive physicalist accounts, mental properties and events hold dependence and determination relations to certain microphysical base properties and events (Moore, 2012, p.324). By my lights, it is hard to see how one could accuse incorporating, coordinated mental-physical overdetermination, even when systemic, of being coincidental or intuitively *peculiar*.

While endorsing systemic overdetermination is non-typical for interactionist dualists in response to the causal overdetermination argument, coordinated overdetermination may also make this response more appealing. The dualist may not posit a dependence relation between microphysical and mental events, but they may envision a tight causal connection between mental and physical causes (Foster, 1991). While I do not intend to fill in the particulars of such a picture now, I have in mind something like the independent coordinated overdetermination picked out in Example 1. For a toy example, imagine “Laura” moving out of the path of a falling piano. Some independent mental event like “deciding to move away from the falling piano” is caused by some neuronal and optic-nerve firing when

Laura's eyes fix on the falling piano. She forms her intention just as her brain sends signals to nerves in her limbs which cause them to move. Both the intention and the neural signals are, on the sketched dualist picture, sufficient causes. Yet, putting aside whether one *accepts* the possibility of such a dualist account, the mental and physical causes have a proximate prior cause themselves.

With NUO, we first reduce the intuitive implausibility of overdetermination in general for non-reductionists of physicalist or dualist inclination. Then, Funkhouser's taxonomy and my coordinated and non-coordinated overdetermination distinction allow us to differentiate the truly peculiar cases of overdetermination from acceptable instances. Neither non-reductive physicalists nor interactionist dualists are compelled to accept a total absence of relation between these mental and physical causes, and both can provide stories which explain these connections and hence, coordination. We need not, in accepting systemic overdetermination, bite the bullet on a 'world of unexplained coincidence' (Bernstein, 2016, p.29) With coordinated systemic overdetermination in far better intuitive standing, I now move to consider some specific objections to systemic overdetermination.

1.3.3 Objecting to Systemic Overdetermination

1.3.3.1 The Worrying Ubiquity Objection

Kim (2005, p.45-8) is concerned that accepting causal overdetermination to "save" mental causation in the causal overdetermination argument risks permitting concerningly widespread overdetermination. To borrow a phrase from Bernstein, accepting systemic causal overdetermination commits us to a 'causally crowded world' (Bernstein, 2016, p.29).

In isolation, this concern is defanged by appeal to NUO, once more. Truly ubiquitous overdetermination makes the world no more crowded by causation than our everyday and scientific practices of causal attribution, which I have asserted undergird NUO, already do. We posit causation at many levels besides the microphysical within social and scientific discourse, as I have discussed. Accepting these notional causes as *actual* causes is in line with our intuitions, as well as our explanatory and predictive scientific practices.

Further, the coordinated and uncoordinated overdetermination distinction provides a framework to somewhat restrict our causal attributions. Accepting systemic overdetermination does not commit us to a host of coincidences, but only to the most plausible sort of overdetermination case.

A closely related objection to ubiquity would be ontological excess (Bernstein, 2016, p.29). It is not merely that widespread overdetermination would seem intuitively peculiar, or that it would present us a crowded picture of causation, but rather that we are positing more than we ought within our ontological framework. I will not dispute the virtue of, nor am I opposed to, avoiding ontological excess. However, it seems to me that such a commitment to parsimony in our metaphysics applies most prominently when all things are equal. Given a choice between two philosophical positions which do approximately equal justice to our intuitions and philosophical commitments, or two scientific theories with approximately equal explanatory and predictive power, concerns about parsimony might encourage us to choose the theory with fewest postulates. I do not take “all things to be equal” here, however.

NUO gives us cause to suppose causation is widespread across levels of reality, and if that assumption is well-founded, then it is reason to embrace some ontological excess. There are other reasons besides NUO; systemic overdetermination allows the non-reductive physicalist to preserve both mental causation and closure (Bernstein, 2016, p.29). Besides which, one may believe for independent reasons that a more elaborated ontology is not necessarily false (ibid, Willard, 2014).

It could be argued that the mental explanation is not excessive, but rather explanatorily *redundant.* A mental explanation is redundant where some physical explanation is explanatorily sufficient and perhaps even more *apt*. This may be for reasons of pragmatism, parsimony, or for the presence of other good-making features one may expect of explanations. If a mental cause is explanatorily redundant, then we might reasonably wonder what causal work it is doing. Fodor argues that higher-level scientific explanations point to interesting and salient features of higher-level properties in a way that a microphysical explanation would invariably fail to do (Fodor, 1974, p.104). What interests us about laws of economics, or psychology, is not something which can be captured by a lower-level description. If this is right then we might decide, at least sometimes, that it is the physical cause which is explanatorily redundant.

However, we should not feel compelled to choose. If we begin from the starting-point assumption of NUO rather than causal singularism, then we assume causation at many levels of reality. On that view, it seems reasonable to suppose we should have explanations for causal relations similarly across these levels. Given this, if there is tension between

explanations, we should not expect mental explanations to be consistently redundant, and we also need not assume the causal singularist view of causation and explanation privileged to the most fundamental level of reality.

1.3.3.2 The Coincidence Objection

I have already covered much of what might be said about the supposed coincidence of systemic overdetermination in 1.3.2. Coordinated, and particularly incorporating, overdetermination, seems evidently non-coincidental.

Coincidence objections rely upon appeal to cases such as Examples 2 and 3, where 'improbable conspiracies and coincidences' (Schaffer, 2003, p.28) seem necessary. Yet the mental-physical case, especially for the non-reductive physicalist, is not relevantly similar to these examples, and I have already demonstrated how we might undercut the intuitive pull of such comparisons.

Two further points are worth making with regard to systemic overdetermination and coincidence. Moore (2012) contrasts independent and dependent overdetermination, similar to my coordinated and uncoordinated division. However, there are distinctions. Dependent overdetermination is separated from independent overdetermination in virtue of one cause being dependent on the other:

'Dependent Overdetermination: some cause P is necessary and sufficient for effect e, but this cause P renders it necessary that event M occurs, while mental causation renders it

necessary that event M causes effect e, so this cause M necessarily occurs, and necessarily causes effect e as well.'(ibid, p.324-5)

Dependent causation only captures mental causes which are dependent on some base physical event, unlike coordinated overdetermination which definitionally includes independent mental causes where they share a relevant causal ancestor. I can agree with Moore (2012, p.325) that cases of dependent overdetermination are non-coincidental and therefore non-worrisome. I differ in taking it that one can *even* preserve non-coincidental overdetermination for the interactionist dualist.

Here, Moore addresses the prospect of independent mental overdeterminers:

'We would have to admit that the world contains massive amounts of overdetermination if we accept it for mental causation [the] amount of coincidence involved here strains credulity. To think that two divergent paths constantly converge is akin to believing that lightning regularly strikes the same neutral venue. It simply defies the odds' (Moore, 2012, p.323).

A similar sentiment is shared by Funkhouser:

'Systematic independent overdetermination, say a parallelism between mental and physical substances, would either be a coincidence on a cosmic scale or would require a divinely arranged pre-established harmony' (Funkhouser, 2002, p.338).

It may not be sufficient to note that independent mental causes can be involved within coordinated overdetermination in virtue of proximate causal ancestry. Moore is asserting that the sheer ubiquity of overdetermination itself, if independent mental causal overdetermination of physical effects were accepted, would represent an astonishing and problematic coincidence.

I will demonstrate that even the central example of Moore's independent overdetermination, independent *mental* overdetermination, is non-problematic by his own lights *given that it is coordinated*. This will serve to establish that coordinated overdetermination best captures the least worrisome set of overdetermination cases. In addition, it will demonstrate the embracing systemic overdetermination is not just an option for non-reductive physicalists.

For Moore, *non-problematic* overdetermination has certain desiderata besides dependence; the 'co-occurrence' (2012, p.325) of these causes must be expected, the mental cause must be necessary for the effect (p.326), and naturally, the physical and mental causes must occur at the same time non-coincidentally (p.325). I shall bracket the last of these desiderata, since it is what I am attempting to establish.

The first, expected co-occurrence, is straightforward. As stated previously, dualists will typically assert some sort of nomically necessary connection, or psycho-physical law, between mental and physical events, in particular bodily and neural events (Foster, 1991). While these mental and physical events may be *independent* in Moore's sense, the causal picture is closer to the firing squad, Example 1, than Example 2. Like Example 1, the mental

and physical events are strictly independent, but coordinated- in the mental causation case, this is a result of prior physical events such as neuronal firings in the brain.

Moore's next desideratum is that the mental cause must be necessary for the effect. In the case of dependent overdetermination, Moore takes this to be satisfied by 'the mental event necessarily [accompanying] the physical cause due to supervenience' (Moore, 2012, p.326). If this is sufficient to establish the mental cause as necessary for the effect, then the overdetermination-endorsing dualist has at least two choices besides accepting a dependence relation.

First, one can once more posit psychophysical laws which explain these correlations and hold with nomic necessity (Hart, 1988, Foster, 1991). These laws often involve the rejection of causal closure, but they need not. Alternatively, one may take there to be an emergence relation between mental and physical properties or events (Robb, 2018). In this case, the mental cause would still necessarily accompany the physical, despite lacking a typical dependence relationship.

While I have intended to show just how non-problematic coordinated overdetermination can be and demonstrated how mental coordinated overdetermination avoids the charge of coincidence even where coincidence is understood as broadly as Moore takes it to be, it is worth recalling the relevant competing assumptions here.

The causal singularist starts from the assumption that causation is restricted to the lowest causal level of physics, while NUO holds that causation, notionally, appears to

overdetermine effects often in light of causes from multiple levels of reality converging on effects. If NUO seems like the right assumption to adopt as a starting-point for causal enquiry, then our presumption ought to be that overdetermination between levels is not peculiar. Beyond this, I have provided reasons for supposing that such overdetermination is non-coincidental. I have shown that non-reductive physicalists and interactionist dualists alike can accommodate non-coincidental mind-body overdetermination within their respective responses to the causal overdetermination argument. NUO provides intuitive support for endorsing systemic overdetermination, while coordinated overdetermination as a category demonstrates the difference between apparently coincidental cases and those which should not trouble us.

1.3.3.3 The Metaphysical Objection

The final objection to systemic overdetermination I intend to address is metaphysical in nature. Bernstein asserts that 'making a difference to the way the effect occurs is important for accounting for the precise causal contribution of each cause' (Bernstein, 2016, p.24). Cases of apparent overdetermination like Example 3 are ‘causally satisfying’ (ibid, p.30), in virtue of our ability to grasp the causal power and contribution of each rock. Bernstein is conceptualizing Example 3 as an instance of joint causation, where each rock is a contributor to the overall effect (p.30). In the mental-physical case, she questions what the causal contribution of the mental event is, absent the causal contribution of the physical cause. Therefore, this is not merely an objection to systemic overdetermination, but to the possibility of overdetermination whatsoever.

Bernstein's causally satisfying story is illustrative of a tendency to assume some sort of "oomph"-y account of causation wherein causes must make a demonstrable causal contribution to their effects independent of other causes, usually through energy transfer or change of momentum. If a physically oomph-y requirement on causes ought to be part of our best causal theory, this is clearly bad news for the prospect of mental-physical causal overdetermination. However, it may be that our only intuitive motivation for such a requirement emerges from an acceptance of causal singularism. If one explicitly or implicitly holds to singularism, then physical causation has a sort of intuitive primacy as a level which informs our causa assumptions. Physical causation does involve energy or momentum change, or some other physical change and it would be at least extremely rare to suppose a cause which does not uniquely contribute to some effect.

There may be some exceptions to this; one is the oxidation of ceric ions by bromate when in the presence of either bromide or sulfuric acid (Lebender & Schneider, 1994, p.7533-4). It seems in this instance that either is sufficient and no additional work is done in the presence of both, at least under certain conditions (ibid). That being said, this is a peculiar case which largely crops up "in the lab", so little may ride on this. I note it only for sake of completeness.

Bernstein's response falls into the same trap I identified for reductive physicalists earlier in the paper. If one wishes to reject the bare metaphysical possibility of mental-physical systemic overdetermination, or even the possibility of instances, then one should not do so by appeal to principles or theories which implicitly accept the core physicalist assumptions. If you already have reason to suspect there is mental causation, this is utterly unpersuasive.

We are, if we accept Bernstein's point here, quite possibly accepting a begging of the question against the individual who does not already accept the conclusion of the overdetermination argument. To emphasise, the claim rests on a sort of causal account which makes sense if one priorities the physical in our understanding of causation, since physical causation does have the sort of character Bernstein describes. However, we have no reason to suppose that all causation functions by adding extra "oomph" to its effect, where there are multiple sufficient causes.

1.4 Conclusion

This paper has aimed to take a novel approach to critique of the causal overdetermination argument. I have shown that the assumptions of the Sparse View, causal singularism in particular, are pivotal to the intuitive pull of the closure and overdetermination premises of the argument.

I first introduced what I term the "Sparse View" causal assumptions, relata, relation, and singularism. In contrast to singularism, I suggested a different starting-point assumption for thinking about causation, which I termed NUO. I then canvassed the closure principle at length, and considered its metaphysical commitments, how it is supported in the litetaure, and what assumptions others like Gibb take it to carry. With Gibb and Lowe's work in particular, I noted that their work involved undermining the relata and relation assumptions, respectively. Following this, there was some elaboration on the ways in which closure might be further tackled via causal models with these problematized assumptions in mind.

While the sections focused on closure principally addressed prior work and related them to my assumptions as identified, the work on overdetermination centered the NUO and causal singularism assumptions. This was because, by my lights, much of the intuitive force of the overdetermination premise falls out of the view, as captured by various cited philosophers, that such common overdetermination is deeply weird and peculiar. Rejecting this is pivotal to rejecting the overdetermination premise overall. I concluded with a number of objections which may be offered against the possibility of ubiquitous overdetermination. Alongside more fine-grained argumentation, I attempted to demonstrate that even these objections involve an acceptance of causal singularism and the narrowed view of causation it seems to support. If singularism can be rejected and displaced with NUO, we are far better placed to welcome overdetermination as an intuitive and expected part of our observations and models. With NUO in place, I also explained how overdetermination may actually be thought to occur, demonstrating that this is not merely a change of intuitions which falls apart when confronted with the need to describe causal relations “out in the world”. This paper, then, has outlined and undermined assumptions which are core to the premises of the overdetermination argument. From the non-reductive physicalist to the cartesian dualist, those who do not accept one or more of the assumptions of the Sparse View have little to fear from the argument from Causal Overdetermination.

Paper 2: The Causal Exclusion Argument, Edwards' Dictum, and Causation at Levels

2.1 Introduction

Kim's exclusion argument aims to expose a problem for non-reductive physicalism. He highlights three propositions which form the core commitments of non-reductive physicalism; these being Supervenience, Irreducibility, and Mental Causation[26]. Here is one formal expression of the Supervenience thesis, from Kim (2003, p.151):

[Supervenience] Mental properties strongly supervene on physical/biological properties. That is, if any system s instantiates a mental property, M, at t, there necessarily exists a physical property, P, such that s instantiates P at t, and necessarily anything instantiating P at any time instantiates M at that time.

Mind-body supervenience can be understood as a minimal physicalist commitment, which holds that mental properties are 'wholly dependent on, and determined by' (Kim, 2005, p.14) physical properties. The fact that there 'necessarily exists a physical property' (2005, p.33) for each mental property indicates a necessary relationship between mental and physical properties, for the supervenience theorist. This connection is not contingent[27], nor a simple matter of covariance (Kim, 2003, p.152). Mental property-instantiations are, then, *existentially dependent* upon certain physical property-instantiations; it is necessarily in

[26] The non-reductive physicalist may have other commitments, but the three listed are those which are currently relevant.

[27] Rather, the existence of a connection is not contingent, even though some supervenience-theorists might contend that the lower-level realiser in such a relation could be something other than matter, in some possible world. Nothing I have to say going forward hinges on this, however.

virtue of the physical property's instantiation that the specified mental property is instantiated (2005, p.34). There can be no change in the mental properties without some corresponding change in the base physical, underlying, properties.

As Kim notes, however, this position does not entail any sort of identity between mental and physical properties. Mental properties are not 'reducible to, and are not identical with, any physical properties' (Kim, 2003, p.152). This Irreducibility commitment, clearly, is what separates non-reductive physicalism from psychophysical reductionism, or type-identity theory. There are motivations for this commitment, primarily the argument from multiple realizability (2003). Ontological reductionism identifies all higher-level properties with base microphysical properties (Kim, 2005, p.23). As Kim states, 'neural bases may differ for different instances of pain, but individual pains must nonetheless reduce to their respective neural/ physical realizers' (ibid, p.25). Type-identity theory is reductionist in ascribing type identities to mental properties and physical properties; for every mental property, there is a corresponding physical property to which that mental property is identical (Kim, 2012, p.175)[28]. Non-reductive physicalism denies the possibility of reductively identifying all mental properties with microphysical properties, usually on grounds of multiple realisability (ibid).

So far, the non-reductive view takes mental properties to be *supervenient*, and *irreducible*. Kim highlights one other proposition held by the non-reductionist physicalist. This is the principle of mental causation (Kim, 2005, p.35). Simply put, mental properties, or more

[28] Type identity theory is distinct from token-identity theory which posits mental-physical identities only between instances, events, or particulars (Kim, 2012, p.174).

precisely, mental property-instantiations[29], must cause other, physical, property-instantiations (Kim, 2005, p.35). This principle can be expressed thusly: 'Some mental events cause physical events' (2005, p.35).

Kim's Exclusion argument sets out to demonstrate that Irreducibility and Supervenience are inconsistent with Mental Causation, in particular that 'physically supervenient and yet irreducible mental properties are causally impotent' (Kim, 2003, p.152), and that any causally efficacious mental properties 'must be [physical] or be reducible to physical properties' (ibid). The specific role each commitment plays in Kim's argument will be made clear in the following section.

The paper moving forward runs as follows. Initially, I introduce the role that "Edwards' Dictum" plays in Kim's argument. Following this, in section 2, I consider what I term the "causal problem" for the Dictum as deployed by Kim. This is, in brief, that the worldview from which Edwards derives the dictum fundamentally precludes causation, among other issues. I also introduce a conception of causation-at-levels and a related difference-making, or commensurability, constraint on causation. Then, building upon this work, I offer a novel conception of the causal model for non-reductive physicalism which need not depend on any sort of downward causation. This model incorporates "behaviours" into the mental level of reality, so understood, allowing us to preserve both mental causation and the causal closure of the physical.

[29] A property-instantiation, or exemplification, at a given time is, for Kim, an event. For an event to be a mental event, it must involve the instantiation of a mental property. I will follow others in the literature here in generally referring to property-instantiations as "properties" for sake of simplicity.

2.2 The Basic Argument

Kim highlights the importance of Edwards' Dictum to the Exclusion argument. The dictum states that 'there is a tension between vertical determination and horizontal causation' (Kim, 2003, p.153). More than *mere* tension, in fact, since vertical determination is taken to preclude, or exclude, horizontal causation (2003, p.153). How are 'vertical determination' and 'horizontal causation' to be understood such that we might observe a tension?

Vertical determination is a synchronic relation between 'macroproperties' (2005, p.36) and their underlying 'microstructure'. Kim offers the example of a bronze lump, possessing certain macro-properties like its colour and shape in virtue of "lower-level"[30] micro-properties. Thinking in terms of micro- and macro-physical properties, macro-physical properties like the structure of a car, body or brain are wholly determined by, and *dependent on* their 'microstructure' (2003, p.153), micro-properties like the configuration of atoms and molecules of certain types. An object like a brain has the macro-properties it has in virtue of its micro-properties and their structure. If two objects have the same microstructure, this will *vertically determine* that they will have qualitatively identical macro-properties.

Horizontal causation, however, is cross-time, or 'past-to-future causal determination' (2003, p.154), in contrast to micro-macro vertical determination. Explanations of how things are can be given in horizontal or vertical terms. We may say that a car is the way it is because it

[30] Here, Kim notes 'the usual practice of picturing micro-macro levels in a vertical array, with the micro underpinning the macro' (2005, p.36)

had those same properties a minute ago (horizontal causation), or because of its assorted micro-properties. In the bronze lump case (2005, p.37), we might consider a single macro-property, such as the lump 'being yellow' (p.37). Just in case the lump possesses the relevant micro-property at a time T, it will necessarily be yellow at that time. This is regardless of anything that happened before T. Further, Kim argues, if it fails to have that micro-property (or some other appropriate micro-property) at that time, no prior events will make it yellow at T (p.37). As Kim points out, Supervenience is a form of vertical determination, whereas Mental Causation is a clear case of horizontal causation[31] (2003). Since vertical determination is sufficient for the occurrence of the mental property, the tension arises insofar as there is no work left for horizontal causation to do in bringing about said mental property. With this thought in mind, the argument will now be offered in more detail.

Take M and M* to be mental property-instantiations, with M causing M*; 'on the supposition that there are cases of mental-to-mental causation' (2005, p.39). M and M* have P and P* respectively as their supervenience bases (Kim, 2003, p.155) or the microphysical properties on which M and M* supervene.

1. [Mental Causation] 'M causes M*' (Kim, 2005. p.39)

2. [Supervenience] M* has P* as its supervenience base (2005)

[31] For example, my mental properties are the way they are as a result of my past mental properties. By contrast, the vertical determination explains current mental properties by appeal to underlying physical micro-properties.

From Edwards' Dictum, there are two ways of explaining the instantiation of M* which are in *apparent* tension, with one seeming to preclude the need for the other[32]. Either M* is "horizontally caused" by M, or M* is vertically determined by P*. Kim takes it that priority must be given to the vertical determination of P*, since M* is necessitated and guaranteed by P* (2005, p.40), regardless of prior mental events.

Since it is clear from earlier discussion that M* would have occurred given the presence of P* *no matter what came before,* the only way of resolving this tension that Kim sees (short of simply abandoning [Mental Causation]) is to suggest that M is involved in the causation of P*:

3. 'M caused M* by causing its supervenience base P* (2005, p.40)

Kim sometimes proposes an explicit "exclusion principle" to motivate (3), instead of relying on the intuitive appeal of the solution he offers, given Edwards' Dictum. More will be said about the potential exclusion principles that could be brought to bear in support of this premise in the next section, but Kim's causal exclusion principle follows thusly (the same principle is deployed in this argument elsewhere, shortly); 'No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination' (p.42). Henceforth, this principle will be called [Exclusion][33].

[32] Whether it is appropriate to understand these different sorts of relation as excluding one another, or even being in tension, will be discussed in the next section.
[33] Kim suggests other forms of exclusion principle elsewhere, such as determinative/generative exclusion.

So after (3), given [Supervenience], mental-to-mental causation appears to lead to mental-to-physical "downward" or inter-level causation, at least insofar as one wishes to preserve mental causation of some sort (2005, p.40). As Kim puts it 'level-bound causal autonomy is inconsistent with supervenience or dependence between the levels' (2005).

Further from [Supervenience];

4. 'M has a physical supervenience base, P' (p.41)

Kim asserts there are strong reasons to suppose that P is a cause of P*, and this is not something which I challenge. From (3), M is taken to be a cause of P*, and the supervenience of M on P seems to qualify P as a cause of P* also, for Kim (p.41).

5. M causes P* and P causes P* (p.41)

Since P-to-M is not a causal relation but a determinative one P->M->P* cannot be thought of as the causal chain here. M is not a link in a *causal* chain between the two physical events. Given [Exclusion], M and P are in competition to be the sufficient cause of P*.

6. [Irreducibility] M=/=P (M is not identical or reducible to P).

This is simply to say that M cannot be a cause of P* in virtue of being identical with the sufficient cause P. That avenue is available to the reductive physicalist, but not to someone who accepts the triad of claims which must be accepted by the non-reductive physicalist

(Irreducibility, Mental Causation, Supervenience). P is a sufficient cause of P*, and from [Exclusion], M cannot therefore be a cause *as well*, unless this is a case of genuine overdetermination (2005, p.42). Thus;

7. P* is not causally overdetermined by M and P (2005, p.42)[34].

Either M or P must be excluded, then, as the cause of P*. One final principle must be introduced here. This is not a principle which will be controversial to the non-reductive physicalist.

[Closure] 'If a physical event has a cause that occurs at t, it has a physical cause that occurs at t' (2005, p.43).

Given closure, we reach our conclusion;

8. P causes P*. M does not.

To briefly recap this argument, M is excluded from directly causing M* by [Exclusion], or as an intuitive result of accepting Edwards' Dictum. To preserve the assumption of mental causation, it is argued, *ex hypothesi*, that M causes the physical base of M*, P*. But we already have a physical cause for P* from Closure, P. Given Exclusion, both M and P cannot be sufficient causes of P*, absent genuine overdetermination. M cannot cause P* in virtue

[34] This is not uncontroversial, and responses to Kim along overdetermination lines will be considered in the next section.

of being part of a causal chain with P, or by being *identical* with P given (6), Irreducibility. Thus, P precludes any causal influence of M on P*. Mental causation, both of mental events, and downward causation of physical events, is threatened.

2.3 A Causal Problem for Edwards' Dictum

The previous section outlined Kim's Exclusion argument. Kim relies on a number of assumptions and principles which I intend to discuss. The first and perhaps most prominent is Edwards' Dictum. This runs, again, as follows 'there is a tension between vertical determination and horizontal causation' (Kim, 2003, p.153). Kim identifies this dictum with a historical position advanced by Jonathan Edwards, the first to observe this kind of tension (2005, p.37). Attention will now be paid to Edwards' original argument and the sort of situation where the purported tension is meant to be observed, in order to establish a dis-analogy between the classic case and the mental-causation case.

Edwards was concerned with the possibility of causation in a world where God was the 'sustaining cause' (2005, p.37). We can imagine a world where 'no persisting objects' exist, since God 'creates [the] entire world *ex nihilo*' (p.37) at every moment. As an aside, in Edwards' imagined world, persistence will always be perdurance, not endurance, strictly speaking, but this technical clarification has no bearing on how we ought to understand causal relations in what follows. For in such a world, there can be cross-time causal relationships between entities, since each time-slice is created anew. The "car-engine running" is not a result of some prior-causal chain, but is vertically determined by the

relevant microstructures and micro-properties "set" in that moment by God. In this case, there is no work for horizontal causation to do nor, indeed, any work that it *can* do, since cross-time causal relations are impossible in such a world. *Pace* Kim, '"horizontal" causation involving created substances is excluded by their "vertical" dependence on God as a sustaining cause of the world at every instant' (2005, p.38).

However, this in itself is a serious problem. The initial picture presented by Kim, from which the Dictum is presumably meant to derive some of its intuitive strength, is of a world where 'we lack a necessary precondition for diachronic causal relations' (Campbell, 2015, p.57). Specifically, there can be no event-causation between moments in the world that Edwards is concerned with. It is not vertical determination which precludes horizontal causation in the imagined world where God is the sustaining cause; rather, horizontal causation is impossible given the 'time-slice'(p.57) model built into that world. If our intuitions about this purported tension are meant to derive from Edwards own example, then the argument is in trouble. We would be unjustified in extending the intuition toward disanalogous cases. Lim takes the disanalogy to extend further, noting that Edwards postulates a world with a single "determiner", God, of all objects and properties, in contrast to a standard physicalist worldview on which vertical determination relations can plausibly include several determiners (Lim, 2014, p.47).

Even granting 'full intuitive force' (Burge, 2007, p.372) in the case of Edwards' Dictum, Burge sees an important misstep here (p.372). Kim equivocates God's sustaining causal import with "vertical determination". God may determine higher-level structures in each time-slice by causing the instantiation of the entire microstructure but *this* does not

establish the tension between determination and causation intuitively. We are left with two competing causal explanations for any given fact, causal "work" from earlier time slices, or the sustaining cause that is God. That our intuitions align strongly with the view that *God* as cause precludes other putative causes within this model does not lend credence to any additional tension between causal and vertical-determination relations.

Aside from these specific problems with the Kimean story offered against the possibility of higher-level causation, same-level causation seems both an intuitive and theoretically significant default position, which demands good reason before abandoning it. It is intuitive because it fits with our everyday habits of causal explanation; we explain economic happenings by reference to earlier economic effects, so too psychological happenings, biological happenings, and so forth. Causal explanations in science appeal to mechanisms and lawlike relations between properties *at levels* as well (Burge, 2007, p.367). We can remain agnostic as to whether such levels of description are ontologically significant (for now), denotive of a real metaphysical hierarchy (List, 2019), but the bare possibility should not be ruled out by fiat. Indeed, it seems that, had P* failed to occur, M* would still have obtained in virtue of some other supervenience-base in some close possible world (Burge, p.373). And in that case, there is no clear reason why M would not still be its cause (p.373). Given this sort of consideration, we may even be justified in taking horizontal causal relations at higher levels to be "prior", or at least on a par, as an explanation for the instantiation of higher-level properties.

As is hopefully clear, determination and causation have different sorts of structure. One is diachronic and the other synchronic, one is necessary and the other contingent. Kim

assumes that both "determine" a given event, but this should not be equivocated. Causal and determinative *explanations* are explanations of different kinds. One Is concerned with how things have come to be as they are, and the other is concerned with why things are as they are at a moment in time (Campbell, 2015, p.60). These explanations are "orthogonal" (Marras, 2007, p.311).

Besides this, causation at levels fits well with a plausible sort of causal principle- namely proportionality[35], 'difference-making' criteria, or causal commensurability (Yablo, 1992, List and Menzies, 2007). This sort of principle demands that causes be "commensurate" or proportionate to the effects they generate; neither too specific, nor too broad. List and Menzies formulate this insight in terms of counterfactuals, taking the relevant properties as the value of variables (Menzies, 2007, p.12). Taking the case of Yablo's pigeon, trained to peck at all red things, and which is observed to peck at a crimson sheet:

"[The] proportional difference-making cause of the pigeon's pecking is not the crimson, but the redness of the target, contrary to what the exclusion principle implies. This is supported by the truth of the counterfactuals:

Target is red → pigeon pecks.

Target is not red → pigeon does not peck.

In contrast, the following counterfactuals are not both true:

Target is crimson → pigeon pecks.

[35] John Gibbons makes a case against proportionality as the right sort of causal principle (Gibbons, 2006, p.96). He argues that proportionality involves separating generalised categories of natural kinds into Thus, these generalisations are useful, perhaps, but fail to track truth.

Target is not crimson → pigeon does not peck" (List and Menzies, 2007, p.9)

Simply put, the redness is what makes a causal difference in the pigeon case, not the crimson, since the pigeon could have failed to see *crimson* specifically, and still pecked. So, contra Edwards' Dictum, microphysical determining bases may be causally sufficient, but prior mental events have the right level of specificity to be candidates for causes of later mental events. Many microphysical configurations may be sufficient for the generation of some mental property M*, but the cause may well still be recognised as M, if M* would not have come about without it in 'relevantly similar' close possible worlds[36] (2007, p.7). Indeed, if M is needed to bring about M* in close possible worlds, and P is not (or P could instead have been P1, P2…Pn), M looks like an ideal candidate for the cause of M*.

List and Menzies avoid Yablo's formulation of commensurability since it is dependent on an understanding of the mental-physical relation as *determination*, which need not be accepted. However, they preserve the central insight that causes ought to make a difference to their effects, in a way that can be tracked through relevant counterfactuals. I should briefly note Bontly's objection to a proportionality principle.

'Suppose that Socrates drinks hemlock and dies. As a matter of fact, he guzzled the hemlock, but let's say he would still have died if he instead sipped it as one would a fine single-malt. His guzzling the hemlock was therefore not required for his death, and so not proportional

[36] Similar possible worlds where M* obtains.

to it. (1992, 276). But does it follow that his guzzling the hemlock did not cause his death?'

(Bontly, 2005, p.339)

I think we can rather straightforwardly say, here, that the *guzzling* did not cause his death. It may have been causally relevant, as many other factors were, but his death was caused by drinking the hemlock and would have occurred in worlds where he did not guzzle it but sipped it, as Bontly himself notes. However, how persuasive one finds my response likely depends on the value afforded to the counterfactual reasoning in favour of proportionality. I put this aside for now, since there seem at least some good reasons to endorse something like a proportional difference-making principle.

So, Edwards' Dictum seems to be a principle we ought to reject, given its lack of intuitive motivation, and its rejection of other reasonable explanatory and causal presuppositions and principles. I now aim to argue that the rejection of the Dictum is not just a problem for the first stage of Kim's argument, but a real threat to the entire Exclusion Argument.

2.4 Same-level Causation, Mental Causation, and Psychological Explanation

Why might the rejection of Edwards' Dictum offer some defence of a non-reductive physicalism? Even those like Block (2003) and Campbell, who take issue with the Dictum, will nevertheless grant mental-to-physical downward causation, from which follows the second stage of Kim's argument. Campbell, for instance, only seeks to preserve same-level

causation more generally. His aim is to avoid the sort of causal drainage that concerns Block, wherein the properties of various sciences (biological properties, to offer one example), *cannot* be said to have genuine causal relations simpliciter, but only in virtue of their microphysical base-properties (whatever those turn out to be). This also concerns Burge, as a threat to the causal relations postulated by these sciences[37].

The reason why this is granted is straightforward; mental property-instantiations are taken to have not just mental, but physical, effects, per a commitment to "mental causation"; specifically the causation of physical property-instances by some mental property-instances. Same-level, or domain-restricted, causation, is not what the non-reductive physicalist is centrally interested in. Perhaps same-level causation can be preserved at other levels without entailing causal drainage, but the non-reductive physicalists' commitment to mental causation is taken to be in conflict with such a model. If mental properties[38] *do not* cause physical properties, then the mental seems worryingly epiphenomenal. Protecting mental-to-mental causation is not a strategy that seems to go anywhere, on its own. I aim to argue that this rests on a faulty assumption, or more specifically, an equivocation between kinds of physical effect or property.

In the previous section, I briefly discussed the notion of causal explanations at "levels", typically understood as levels of scientific inquiry and description (List, 2019, p.854). In chemistry, for instance, we postulate chemical properties in interaction with each other, governed by laws indexed to that level. While causal explanation may not be a perfect guide

[37] The causal relations at levels besides microphysics might well be disputed, but we should not dictate or circumscribe the causal relations of special sciences without good reason.
[38] Again reading "properties" as shorthand for their instantiations, for simplicity.

to real causal relations, it is a reasonable starting point. Aside from this, the potential for multiple realisability suggests a metaphysical picture wherein the properties of special sciences which cannot be reduced or identified with microphysical properties are instead grounded in, or supervenient upon them, with a hierarchal or levelled structure.

We sometimes assume a good intuitive understanding of the properties picked out by these sciences. In psychology, the properties that feature in explanations are *psychological properties*. Since psychology is concerned with the mind, and since psychologists spend much time discussing belief, desire, motivational states, and so forth, it might be assumed that the psychological properties are exhausted by the *mental* properties. But this is not the case, and it not only fails to do justice to psychological explanation, but is, perhaps, responsible for substantial and unnecessary confusion vis-à-vis the exclusion argument and the minimal commitments of the non-reductive physicalist.

Psychology is not only descriptive and predictive with regard to mental properties, but behavioural properties. Not that long ago, behavioural properties were very nearly the *sole* object of inquiry (all the way to BF Skinner). And while this has changed, behaviour, in terms of effects and causes relating to mental states, remains integral to psychology. So too language, and tool-wielding phenomena, to flesh out the picture further. Such phenomena clearly feature in psychological explanations. I also see no reason to suppose the same phenomena cannot be realised by various distinct microphysical structures, just as mental properties are purported to be, on the basis of multiple realisability[39].

[39] Multiple Realisability (MR, henceforth) is not uncontentious, but I will assume its possibility for now given that the non-reductive physicalist is already committed to it. If MR is false, then the physicalist has compelling independent reason to accept an identity theory.

When I talk of behaviours, I mean any intentional action, movement, or locution on the part of a conscious agent, or actor. A given behaviour might be realised by (1) me, (2) my cyborg duplicate, or (3) Clone-me with variant neurological and physiological structures, holding all other mental properties constant. Assuming the same motivational structures, emotions, personalities, and so forth, prior psychological properties seem to be the best candidates from (1)- (3) for commensurate, or difference-making causes. The behaviour, then, accepting the difference-making principle, should best be understood as having my prior psychological property as its cause, not the microphysical base property of that psychological property[40]. If my cyborg duplicate is compelled to buy Brand Y Coffee as a result of some desire, guided by some psychologically manipulative advertising, then so too would I have been. The psychological explanatory story remains the same and extends to the same behaviours. Behaviours thus seem to function as difference-making variables within psychological causal *explanations*[41]. To go even further, these behaviours will often motivate future psychological states, like my beliefs- confabulatory or not- about my prior behaviours, and later psychological consideration of those behaviours. Psychological states and behaviours seem rather appropriately understood as forming parts of causal "nets", interconnected across time. Any one behaviour could, in some close possible world, be differently realised, without seeming to interfere with the emergence of the "same" mental

[40] My discussion of irreducible psychological and behavioural properties interacting causally intra-level may bring to mind a Davidsonian view (1970) on which psychological and physical explanations conflict; the lack of exceptionless laws in psychology ensures it cannot be regarded as a science akin to physics. I am not endorsing this view. For reasons outside the scope of this paper, I am not convinced of Davidson's view on the implausibility of exceptionless psychological laws (Davidson, 1970, p.108).

[41] While causal explanation and causation might reasonably come apart, I will take it to be a reasonable principle that our causal explanations ought to guide our understanding of causation, all things being equal. I will further take it that the burden of argument is upon those who wish to argue that our quotidian or scientific causal explanations are getting something wrong metaphysically.

properties down the line. This seems at least plausible if we take multiple realisability seriously, and I take it that those endorsing non-reductive physicalism already have reasons to do so.

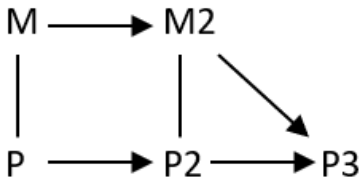
So behaviours are not best understood as merely macro-physical property-instantiations, but as non-reducible "physical" properties which supervene, or are grounded in, base-physical property-instantiations[42]. This brings us to the equivocation I mentioned earlier. It is already granted (by the non-reductive physicalist) that mental properties are irreducible *and* in-some-sense physical. Given, then, that room already exists for the irreducibly physical, we can suppose that "Mental Causation" consists in mental properties being causally related to physical, yet irreducible, behavioural phenomena. This, I think, does not just do justice to scientific psychological explanation, which postulates relations *at that level*, but it also yields an intuitive model for human causal interactions. My mental properties are ordinarily not directed at changing microphysical properties[43], even supposing that they have indirect causal effects on the microphysical. These entities seem far too fine-grained, and my actions too coarse-grained, to suppose my acts are direct and proximate causes of microphysical properties. Phenomenologically, I interact with my own body, with tools, and with objects insofar as they 'feature' in my mental world.

[42] Gibbons offers a similar view in regards to behaviours and causal sensitivity, which he deploys in place of the proportionality principle I endorse. However, we are alike in regarding psychological states and behavioural states as often the right kind of difference-making cause: 'Human pain or pain realized by physical state P is no more a psychological kind than African emerald is a geological kind. The difference between human and Martian pain might make some kind of difference, but it doesn't make a mental difference' (Gibbons, 2006, p.102)

[43] Mental events seem to poorly commensurate with microphysical effects also, pace Yablo (1992). That said, those occasions where microphysical properties do seem to be in causal interaction with our mental properties will be discussed later; I shall explain why such instances present no problem for the view under consideration.

Thus;

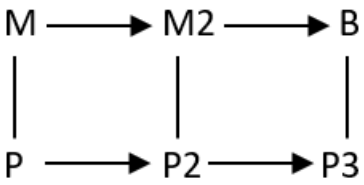
Figure 1- The Standard Model



Take vertical lines to be supervenience or grounding (vertical determination in Kim's nomenclature), depending on ones' preferred vertical relation. Nothing I have to say here rests on this. Arrows denote causal relations.

As mentioned, on this model, 'preserving' mental-to-mental causation does little, dialectically speaking, for the non-reductive physicalist who wishes to preserve mental-to-physical causation. Downward causation must still be invoked, it seems, and consequently the mental cause is excluded by a sufficient physical cause.

Figure 2- Behaviour Incorporated



B= Behavioural property/property indexed to psychological "level"

On Fig. 2, the mental does not downwardly cause any microphysical effect. I do not rule this out as a possibility, just as one should not rule out the downward causal efficacy of chemical or biological properties. But it need not be regarded as a necessary or ubiquitous feature of

causal relations for the mental. To reject overdetermination concerns we do not *need* to deny all downward causation, only ubiquitous downward causation. And "causing behaviours" captures the essence of a belief in mental causation. One need not suppose that mental causation entails direct causation of specific microphysical effects. Mental causes are causes within the domain, or at the level, of psychology.

Further, nothing proposed here violates Kim's *causal* exclusion principle, in general. Rather, I am offering a picture of mental-to-physical causation which does not generate a causal exclusionary worry, nor which yields persistent causal overdetermination. This is because of the aforementioned distinction between microphysical properties, and the macrophysical and behavioural properties which mental properties *actually* enter into causal relations with. It *does* violate Kim's principle of determinative/generative exclusion (2005, p.17), inasmuch as there still exist two determining properties for the instantiation of B, but there is little reason to accept this broader principle in the first place.

> 'Principle of determinative/generative exclusion: If the occurrence of an event e, or an instantiation of a property P, is determined/generated by an event c—causally or otherwise—then e's occurrence is not determined/generated by any event wholly distinct from or independent of c—unless this is a genuine case of overdetermination' (Kim, 2005, p.17)

B in Fig. 2 is clearly generatively overdetermined, since both M2 and the supervenience-base P3 are sufficient for its generation. However, once more, determination and causation are different kinds of relation, with different structures and different roles.

Determinative/generative exclusion is clearly motivated by Edwards' Dictum, and ought to be rejected for the same reasons. We can and should permit more than one "generative" relation to any given property-instantiation, *even if* we do not wish to grant more than one sufficient cause for any instantiation[44]. I see no independent reason besides the dictum for accepting *this* principle. It seems entirely plausible and non-worrisome that a given event may be generatively overdetermined.

In fact, I think events may even be often generatively caused or determined "out in the wild". For an example from human biology, someone may develop a hormonal imbalance which leads to a host of other issues. We might suppose this is low-testosterone. That same condition *could* be generated upwardly by either insufficient nutrient intake limiting testosterone production, or by some gonadal issue preventing production. Since these issues often interrelate in clinical cases, we can reasonably expect them to often come together. Further, they might bottleneck testosterone to just the same level jointly as either might alone.

There may be several plausible candidates for generative overdetermination, in fact. Imagine a person with light-filtering shades which block out any amount of light above some threshold, or some window pane constructed for the same task. In either case, we could imagine multiple external sources of light, all of them sufficient for some event like "such and such level of light being visible". Yet, any one such source would be singularly sufficient, also.

[44] Outside cases of "genuine overdetermination".

2.5 Concerns with this Picture

2.5.1 Cross-Level Causation and Overdetermination

One sort of objection which might be raised against the model I am presenting, which attempts to preserve same-level causation for the psychological, is the possibility of cross-level causation. I do not think this *is* a problem with the incorporated-behaviour model, but a discussion of cross-level causation will hopefully be illuminating nonetheless, and should put to rest such worries.

What sort of cross-level causation *could* be understood as problematic here? Examples which bridge the gap of direct causation between discrete levels; such as human using a computer to control a laser, the beam of which precisely affects some specific microphysical properties, or a Geiger counter, which brings some microphysical properties (through that measurement apparatus) to the attention of human agents, thereby having effects on their psychology. There are also instances between *other* levels, such as economic properties being affected by the psychological properties of some individual- say, the chair of a large bank, or the finance minister of a country. These cases, amongst many others, would be a problem for someone committed to a strict delineation of causation at levels. However, this is not an aim, nor an implicit principle, of the position I am offering. I am entirely happy to suppose that downward, or upwards, causation might sometimes occur, as an overdetermining cause.

Recall Kim's Exclusion principle: 'No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination' (Kim, 2005, p.42). I take there to be reasons for doubting the plausibility of this principle in itself, which will be discussed shortly but *even* this restrictive principle, and Kim's further discussion, permit that cases of overdetermination will *sometimes* occur. The worry is only with a view that makes such overdetermination *ubiquitous*. Certainly, we will observe overdetermination in these somewhat peculiar constructed cases, but this is not an indication of a widespread phenomenon. And not only can we accept that these (and perhaps other) cases of overdetermination are non-problematic for, as Funkhouser notes, 'not all overdetermination is created equal' (2002, p.336). We can turn to a distinction made, in slightly different ways, by both Moore (2012, p.324-325) and himself (Funkhouser, 2002) between independent and dependent, or 'incorporating' (2002, p.337), causal overdetermination to clarify why such instances should be regarded as less worrisome than the the cases of overdetermination that Kim proffers.

The classic case of 'independent causal overdetermination' (Funkhouser, 2002, p.337) is that of a man being killed with multiple bullets, from different guns, at the same time[45]. We imagine two (or more) bullets hitting the target's heart, such that either would be sufficient to cause his death. What are the salient constituents of such a case? First, there are multiple 'sufficient and distinct causes for the same effect' (Funkhouser, 2002, p.335)., *ex hypothesi*.

[45] I have discussed some issues with this example, and independent causal overdetermination quite generally, in my earlier paper on the Causal Overdetermination argument. For present purposes, I assume that independent causal overdetermination is a conceptually coherent phenomenon, which can occur in our world, and which, *ceteris paribus*, reasonable accounts of causation ought not to rule out. Indeed, both Kim and others in the Exclusion Problem debate seem, justifiably, unwilling to rule out the *possibility* of independent causal overdetermination, even if they suppose it is an irregular occurrence, or the sort of coincidence which should not play a role in systematic explanations.

Not only do these causes implicate distinct efficacious *properties*, but they emerge from independent causal pathways or generative mechanisms (p.338). In the given case, these pathways are distinct guns being fired by independently acting agents.

Independent causal overdetermination seems, clearly enough, to differ from that being posited in cases of mental-to-microphysical causation, or in other cases of inter-level causation isomorphic to Figure 2. As Kim himself notes, the 'usual notion of overdetermination involves two or more separate and independent causal chains intersecting at a common effect' (2005, p.48). The intuitive *concern* with this independent overdetermination is that, as part of a system, or explanation, it would appear to be a remarkable 'coincidence' (Melnyk, 2003). However, the causal overdetermination of some microphysical event is by both a preceding microphysical cause, and a mental cause which is vertically realised *by that same microphysical cause*, so this is not a case of two wholly distinct causal pathways or generative mechanisms. Rather, if one could intervene on the preceding microphysical cause, then neither the mental cause nor the physical effect would come about.

To illustrate precisely what is going on here, suppose P1 is the microphysical base of M1, and that M1 and P1 (both at time T1), cause P2, at time T2. P1, on Kim's own understanding, is causally responsible for P2, and generatively responsible for *both* P2 and M1. Therefore, both causes are borne of the same generative mechanism. And this causal picture is closer to Moore's dependent overdetermination, or Funkhouser's incorporating causal overdetermination, than to worrisome independent overdetermination.

Both dependent and incorporating overdetermination involve, again, distinct causally efficacious properties, but in this case they are 'predicated of the same mechanism' (Funkhouser, 2002, p.340). Dependent overdetermination picks out any case where some cause C is both necessary and sufficient for some effect E, while also necessitating that some further event N occurs, which is also sufficient to cause the effect E. Incorporating overdetermination shares this characteristic, but Funkhouser takes the overdetermining causes centrally to "work through" the same mechanism, as a result of the sort of realisation or necessitation relation which Moore also points to. The realisation relation between the microphysical base property and the "macro" property, means that the micro-cause is incorporated in the macro-cause (Funkhouser, 2002, p.340). Key to both understandings, and what distinguishes these cases from independent causal overdetermination, is that the overdetermining causes share some proximate generative event[46] in common, prior to these causes overdetermining some effect. Moore offers an example to demonstrate how two distinct causes might emerge from the same mechanism.

> 'Imagine we reconstruct our initial firing squad example in the following way: a murderer is sentenced to death by fire. A fire is lit and the toxic smoke kills the man along with the searing heat. In this case, the smoke and the flame both kill the man together […] The smoke is dependent on the flame, and this intimate relationship makes their co-occurrence very much expected.' (Moore, 2012, p.325)

[46] This event may well be one of the overdetermining causes.

In such an example, it is clearly non-coincidental that the two causes "ride together" in leading to the same effect, and this quite neatly maps to the above illustration of the psycho-physical causation case. Dependent overdetermination is not an entirely uncontroversial notion; I discuss some further issues in another paper. However, it seems intuitive that the psycho-physical case, as outlined, is less "co-incidental" than the initial "execution" example. We can point to a clear connection between these causes, as they emerge from the *same generative effect*- an entirely natural feature of this picture. Their convergence on a given effect also becomes non-coincidental, in those cases where it does occur. Furthermore, the *relative* paucity of overdetermination which might be assumed on the "Behaviour Incorporated" model gives us less reason to worry about occasional overdetermination *in the first place*.

However, the threat of overdetermination is not the only worry that might be raised with regards to this model.

2.5.2 The "Pre-Established Harmony Between Levels"

Kim has this to say about the apparent "harmony" between levels of reality;

> 'It would be nice if we could embrace causation at many levels, including the psychological, the biological, and so on, and also cross-level causation, both downward and upward, all of them coexisting in harmony.' (Kim, 2005, p.64)

One worry the physicalist might have is that, without identity, it seems *coincidental* that apparent psychological chains of cause and effect, for instance, seem to run in lock-step with lower level microphysical chains of causation. I take this to be a sort of intuitive concern, rather than a formal argument against non-reductive physicalism. That said, I think this can be concern can be readily quelled.

As already mentioned, the relation between mental and physical properties, and the effects of mental and physical causes, is not best understood as some manner of remarkable coincidence. Some mental property at time t1 is vertically determined by some physical property at the same time, and these cause, respectively, some mental property at t2 and some physical property at t2. They flow from the same point, as it were. Why then, it may be asked, does that mental property at t1 not cause some other mental property, not vertically determined by the physical property at t2?

Two points are worth noting here. Firstly, the non-reductive position holds that mental properties will always supervene on *some* physical property; we can take them as wholly dependent upon physical properties. Building on this, per multiple realisability, we can further note that the mental property at t2 could have supervened on some *other* physical property and It just so happens that it was that physical property in our world. This is not at all miraculous; if we could "look inside" other close nomologically possible worlds, we might find other physical properties as the vertical determiners of that same mental property.

Building upon these points, it might be argued that the entire "coincidental" characterisation misses the metaphorical, or metaphysical, forest for the trees. If our

starting-point here is defining the mental-physical relation and the apparent relation between mental causes and physical effects, we should begin by taking the world as it is. We already recognise, as part of our common-sense or quotidian picture, an intimate relation between the mental and physical. A grounds of the mental causation premise itself is the powerful intuition that there are mental causes of physical effects, and our best theories should just do justice to that. If the objection is that ontologically distinct mental causes seem to coincidentally converge on physical effects, it does not seem apparent *why* we should be surprised by this convergence. There does not seem reason to credit that dependent and supervenient mental causes should be surprising in generating these physical effects, when we could equally view such as an obvious outcome of *intimately* but non-causally related properties exercising their causal powers. Indeed, on non-reductive physicalism, these mental events do not correlate at random with their base-physical events, but are in-some-sense generated by them

Further work Is required regarding what my position could say about the demarcation between levels itself, and in particular how we are to demarcate them. My dialectical opponent might be concerned that I am not just attributing mental events to one level of reality but behaviour too- questions that are confronted by non-reductionists in general about levels of reality (List, 2019) may therefore be all the more salient for one focussed on my proposed model. As List (2019) notes, levels are delineated in a number of fashions; explanatory value or featuring in some scientific-model, informational complexity, subsumption under laws, proportionality, and more (ibid). Dealing with this issue would require its own paper, but I think something can be said, at least, about why this problem is for the many and is not especially acute for myself. Non-reductive physicalists in general

accept that the psychological or mental level is distinct, though naturally dependent on and derivative from, the microphysical level. My only addendum to this is behaviours, and the question then seems not to be whether we can demarcate levels in general or how to do so, but whether behaviours are rightly demarcated at the same level as beliefs, intentions and so forth.

Prima-facie, behaviours fit at the relevant level on most reasonable methods of delineating levels. They *feature* in explanations at the level of psychology, they are proximate causes for reasons I have discussed, and they seem to function at a similar level of complexity to mental or psychological events which seem to cause them.

2.6 Conclusion

This paper began with a summary of the exclusion argument against non-reductive physicalism. The argument concludes that mental events are excluded from being causes in light of causal closure, which guarantees that there is a physical cause for each event, and the exclusion of prior mental events from being causes of latter mental events where there is already a base-physical event which is generatively and synchronously responsible for bringing about that mental event.

The latter principle is defined as Edwards' Dictum, which holds there to be a tension between vertical determination and horizontal causation. I first set about undermining the intuitive force of the Dictum, to allow that one could have both vertical determination of mental events by physical events without apparent tension, before building a model on

which irreducible mental properties can be causal. The concern from that point, as I saw it, was that mental events determined by physical base events still seem to overdetermine the later microphysical effects which their physical bases cause, on causal closure. I proposed that mental causes need not be taken to compete for causal influence with these underlying physical events, but rather can be understood as causing macro-physical events, behaviours and similar level phenomena. Spelling out this view, I appealed to difference-making and commensurability causal principles, due to Yablo and List and Menzies. Finally, I defended this expanded view of mental causation for non-reductive physicalism against certain objections.

If all this is right, it could nevertheless be permitted that mental causes are sometimes attributable to later physical effects, but my position would still undercut the sort of overdetermination objection which suggested this was a concern in the first place. This is because this overdetermination would not be ubiquitous and mental causes would typically have effects at their own level. We are left with a non-reductive physicalism that need not be concerned about epiphenomenalism for mental events, and which can tell an intuitive and comprehensive causal story about interactions between mind, body, and world.

Paper 3: Multiple Realisability, Irreducibility, and the Commitments of Physicalism

3.1 Introduction

'I take physicalism to be the view that every real, concrete phenomenon in the universe is … physical' (Strawson, 2006, p.3)

'Physicalism is the view that the world is the way it is, because the physical world is the way it is' (Kallestrup, 2006, 459)

This paper argues for the following three propositions:

(1) Current formulations of non-reductive physicalism fail, or are liable to fail.

(2) Their failure stems from a common cause, specifically the sort of multiple realisability concern which motivate many physicalists to first adopt these positions.

(3) Loosening *one* commitment of non-reductive physicalism allows us to develop a more promising alternative formulation of the view.

I take for granted that there are both good reasons to be a physicalist, and good reasons for physicalists to be concerned with multiple realisability. Given this, and the failure of current positions as I will outline, we are in need of a new formulation of non-reductive physicalism. I will offer two promising options toward the end of this paper. We must begin, however, by

laying some groundwork with regards to physicalism in general, and non-reductive physicalism in particular, so that we can understand just what commitments any proposed formulation of non-reductive physicalism ought to satisfy.

At first blush, physicalism is simply the view that everything, at least everything instantiated in the world, is physical. *Reductive* physicalism holds that all "higher level" properties are physical in virtue of being identical with some physical property. The classic illustration of this, regarding the mental, would be "Pain= C-fibers firing", or "every pain is identical to some pattern of neuronal firing". Non-reductive physicalism holds that all properties are physical insofar as they depend[47] for their existence, in some appropriate manner to be explored later, upon micro- or macro-physical properties. Non-reductive physicalism (henceforth NRP), however, takes mental properties to be *irreducible* to properties of the brain. In other words, a denial of the view that 'mental properties are identical with physical properties' (Crane, 2001, p.2). This thesis can be rendered as follows, to quote Kim:

'[Irreducibility]: Mental properties are not reducible to, and are not identical with, physical properties' (2003, p.152).

Some NRPists will not *explicitly* adopt "Irreducibility" but will rather appeal to something like "Distinctness": 'Mental properties are distinct from physical properties' (Nagel, 1986, p.51). This paper will not consider potential differences between these theses. What is

[47] This usage of "dependent" is not meant to denote any philosophical term of art; for now I wish to remain neutral between relations identified by self-described physicalists to capture the relation between the base physical properties and all other properties. I discuss potential formulations of this relationship in Section 1. I also note that whatever this relation is meant to be, the non-reductionist will want to rule down strongly emergent properties; properties which have downward causal effects upon base physical properties.

important, for present purposes, is that Distinctness aligns with Irreducibility in rejecting the existence of an identity relation between brain or microphysical properties, and mental properties. Therefore, I will continue to use "Irreducibility" for simplicity's sake, to pick out a key thesis of NRP, securing its "non-reductive" aspect.

Yet many non-physicalists will also endorse Irreducibility. Something else is needed to secure the "physicalist" side of NRP. I will call this thesis "Dependence" [48].

[Dependence]: All instantiated properties (which are not identical with physical properties), including mental properties, are entirely dependent upon instantiated physical properties.

As an example, the chair may entirely depend upon its physical, wooden, constituent parts for its existence; without the relevant dependence-base properties, the chair would not exist. While Dependence is vital for non-reductive physicalism, more may still be needed. Dependence may fail to preclude apparently non-physicalist "strong emergence"; the view that there are causally efficacious, irreducible, non-physical properties which nevertheless arise from physical properties. Further, Moorean non-physical moral properties would not be ruled out by the Dependence thesis.

Moore takes moral properties to be non-natural yet supervenient upon natural properties (Moore, 1922, p.286). Though he did not use the language of supervenience at the time, this has been the standard understanding of Moorean moral properties, based upon his

[48] This is complicated by the variety of relations used to describe this dependence of the mental upon the physical in the literature; Composition, necessitation, realisation, to name a few. This discussion is put aside for later.

characterisation of the relation between natural and moral properties (Moore, p.286, Zangwill, 2005, p.11). Such properties would be instantiated, and dependent, yet non-natural and non-physical. One can dispute whether Moore succeeds in carving out a non-physical yet dependent space for moral properties ontologically but this is far outside the scope of this paper; it suffices to say that some philosophers have reasons to think such a space exists, and we ought to refine the physicalist thesis such that it is incompatible with such a clearly non-physicalist possibility[49]. Indeed, even if one is not motivated to adopt a Moorean view of moral properties, then necessary properties such as my property of being self-identical provide another example of supervenient yet non-natural properties.

There are alternative theses to Dependence, chiefly supervenience. In formulating the physicalist thesis, Kim states the following[50]:

> 'Mind-body supervenience can usefully be thought of as defining minimal physicalism—that is, it is a shared minimum commitment of all positions that are properly called physicalist, though it may not be all that physicalism requires' (Kim, 2005, p.13)

NRP is often framed in terms of a commitment to global supervenience (Campbell, 2011, p.37). While there are various distinct supervenience relations laid out in the literature, my second paper deploys the following definition (due to Kim (2003, p.151):

[49] With this concern registered, I think it reasonable to say that if formulations I present later fail to do so, they would be no worse off than most canonical formulations of NRP in this regard.
[50] This, notably, also remains compatible with Moorean moral properties.

[Supervenience] ‘Mental properties strongly supervene on physical/biological properties. That is, if any system s instantiates a mental property, M, at t, there necessarily exists a physical property, P, such that s instantiates P at t, and necessarily anything instantiating P at any time instantiates M at that time’.

There, I directly engaged with an argument of Kim’s, so it was appropriate to focus on Supervenience-NRP. Here, however, I intend to remain neutral between the ways in which NRP may be precisely cashed out. Therefore, for now, we will continue to refer to Irreducibility and *Dependence*, which even supervenience-physicalists will necessarily accept. As Kim himself notes, regarding the core notion supervenience is meant to pick out; ‘It will suffice to understand [Supervenience] as the claim that what happens in our mental life is wholly dependent on, and determined by, what happens with our bodily processes’ (Kim, 2005, p.14). Contrariwise, NRPists such as Melnyk are explicit in noting that supervenience does not feature in their formulation of NRP (Melnyk, 1994, p.224). Any self-identifying NRPists who do not accept, at least, both Irreducibility and Dependence are not the focus of the dialectic in my paper[51]. I am concerned with NRP as picked out by these theses, and later constraints to be elaborated, which I take mainstream NRP to be committed to in virtue of the intuitive NRP picture.

Some physicalists may wish to make comments about *uninstantiated* mental properties as well, though this would seem to go beyond a minimal physicalist thesis and is not within the purview of this paper. Dependence captures that it will necessarily be the case that every

[51] I am unaware of any philosopher with a picture of non-reductive physicalism which would reject one or both of Irreducibility or Dependence, but “just in case”, this paper examines a particular widely-shared basic notion of NRP, not any dissenting position which may be available.

mental property *instantiation* is in virtue of some physical property being instantiated (Kim, 2005, p.34), and that each mental property is 'wholly dependent' (p.14) upon some physical property.

The move to NRP from reductionism is often motivated by arguments from, or concerns about, Multiple Realisability (henceforth MR). A first gloss on MR is that mental properties are (if MR examples hold up) capable of being realised by more than one underlying physical property. Summarising the received wisdom regarding MR, LePore and Loewer state the following:

> '[psychological] properties (including content properties) are not identical to neurophysiological or other physical properties. The relationship between psychological and neurophysiological properties is that the latter realize the former. Furthermore, a single psychological property might (in the sense of conceptual possibility) be realized by a large number, perhaps an infinitely many, of different physical properties and even by non-physical properties' (1989, p.179)

For some examples, some other animal with distinct neural architecture, or an alien with a silicon brain, or an AI, might turn out to be capable of feeling the same pain as a human, without the firing of C-fibers[52]. Since no single neural property is then correlated with the relevant mental property, no brain property can be identical with said mental property, and reductive or type-identity physicalism is undermined.

[52] This paper will consider certain objections to MR but, as noted, does not attempt to argue for the truth of MR, and is strictly concerned with what follows if one accepts that it entails non-identity.

Although I will say more about MR later in this paper, I do not intend to argue *for* it. If one feels any of the arguments against the possibility of multiple realisability hold water, they may "get off the bus" early here. For now, it is only noted that MR motivates NRPists to adopt some view which holds to Irreducibility and Dependence in place of positing an identity relation between the mental and physical.

There is a certain *intuitive picture*[53] which falls out of Dependence, for the NRPist, summed up here by Kim:

> 'the idea that mental properties are "realized" or "implemented" by physical properties carries with it a certain ontological picture of mental properties as *derivative and dependent*. There is the suggestion that when we look at concrete reality there is nothing over and beyond instantiations of physical properties and relations' (1992, p.6, emphasis mine).

I take NRP to be an umbrella term for a number of well-specified "physicalist" positions which accept Irreducibility, and Dependence. To list but a few, there is Realisation, Supervenience, Constitution, and Token-Identity[54].

[53] The intuitive picture of physicalism, outside specific commitments or theory, plays a significant role in willingness to adopt particular formulations, as I take it. Thus, it is worth keeping in mind going forward.
[54] These formulations are not necessarily mutually exclusive; one might, and many do, couple token-identity and some form of supervenience, for instance.

Holders of such views are likely to accept this intuitive picture and provide a metaphysical accounting for how mental properties are dependent upon physical properties by demonstrating that mental properties are *nothing over and beyond* (often phrased as "nothing over and above") the physical properties. To avoid positing additional properties atop the physical properties, to position mental properties as somehow "derivative", is to remain as *ontologically innocent* as possible, while maintaining Irreducibility. This move appears *prima facie* reasonable. If the mental properties turn out to be something "beyond" or "above" the physical properties, our picture begins to look more like a dualism. It also fails to be "ontologically innocent" in a way that the NRPist would want. NRP aims to be innocent in its ontological picture, not stipulating more entities or properties or "additions to being". The mental must "rely" on the physical for its existence. See also Crane:

> 'For on [NRP], the idea that X supervenes on Y is intended to express the idea that X is nothing "over and above" Y. The idea that everything supervenes on the physical, for instance, is intended to express the necessary determination of everything by the physical. The nonphysical is nothing "over and above" the physical because, given the physical facts in our world, any world with those facts (and no others) must contain the same nonphysical facts' (Crane, 2001, p.3)

Thus, out of Dependence and a desire for ontological innocence, we see that a theoretical commitment is seemingly generated for any position to be both non-reductive and physicalist; namely that mental properties be nothing over and above the physical properties. I say more about this "nothing over and Aboveness", henceforth NOAA, later.

NRP runs into several issues within the literature, not least of which is the "Exclusion Problem", which I cover at length in a my work on causal exclusion[55]. However, I will argue that NRP faces another less-acknowledged problem; for any well-specified NRP position, multiple realisability threatens to undermine the metaphysical relations suggested between mental and physical properties. In brief, one can construct MR-style arguments which problematise any of these positions. The denial of MR cannot be the answer since that ground is already ceded by NRP. I take there to be a solution up for grabs for NRP, and to explicate the solution I introduce a division between "thin" and "thick" forms of non-identity, where "thin" non-identity is endorsed by current forms of NRP, and "thick" non-identity is what I argue for. "Thick" non-identity, once spelled out, is likely to seem intuitively unappealing to the NRPist, but I intend to offer reasons to think a certain "thick" view can fit neatly with the intuitive NRP picture, and to show that it is modally set apart from any sort of dualism while conforming to the modal picture inherent in NRP. Further, the adoption of this view will carry other theoretic benefits, to be elaborated.

I will now briefly reiterate and elaborate the shape of my argument over the next two paragraphs, before stating how the paper will run, section by section. This argument introduces a number of novel "moving parts", so the following should clarify the stakes, target position, problem, and my eventual solution. Physicalism is something like the view that every instantiated property is a physical property. Physicalism was once most commonly understood as a thesis of type-identity; the idea that every property is identical

[55] Briefly, if mental properties and events occupy the sort of role described by NRP- non-identical to the causal physical properties but also not properties with independent causal powers- then mental properties and events seem excluded from being causes since the physical properties are sufficient for later events. These properties also cannot be additional sufficient causes if one accepts, as many do, some sort of Exclusion principle on systemic multiple sufficient causes. This is discussed further in my second paper.

to some physical property. Some physicalists, NRPists, are motivated to abandon this identity between mental and physical properties by MR, which seems to show a many-to-one relation between mental and physical properties which is inconsistent with their identity. However, these NRPists want to maintain that while mental properties are not identical to physical properties, they are wholly dependent upon physical properties and are, therefore, really nothing over and above them. The commitment to mental-physical non-identity is Irreducibility, of course, and "wholly dependent" here picks out Dependence, forming a conjunction of propositions to which NRP is committed.

I shall describe a number of NRP formulations which share commitments to Dependence and consequently nothing-over-and-aboveness, and thereafter demonstrate that they fall into problems relating to MR. Spelling out these positions will allow us to see what they share, which is attempting to satisfy nothing-over-and-aboveness through the adoption of "thin non-identity", a concept elaborated later in the paper. Briefly, however, a position which suggests a mere thin non-identity between mental and physical properties is one which either posits an implicit identity relation, or somehow deploys identity elsewhere in its framework. Moving away from such formulations in light of apparent failure, I suggest modifying or loosening nothing-over-and-aboveness in a way that permits thick, rather than thin, non-identity between mental and physical properties while still satisfying the spirit of Dependence. With the class of problem picked out through demonstration through these canonical NRP formulations, and the root of the problem identified, I present this thick non-identity as a way out for NRP. I sketch two routes one might take after embracing thick non-identity, W-grounding and the Causal View. Finally, I consider additional theoretic benefits of adopting either such view. These benefits may serve to soften some theoretical sting for

the target audience of my paper; those committed to Irreducibility and Dependence who have taken it that they are additionally committed on Dependence grounds to the notion that mental properties are nothing over and above physical properties. Nothing-over-and-aboveness is not required to secure Dependence, as I will argue, hence de-prioritising nothing-over-and-aboveness and the NRPist can maintain Dependence on a view where mental properties are truly, thickly non-identical to physical properties, and caused by them in a manner I will explain.

In what follows, I first elaborate on some of the considerations above. I begin, in my first section, with further analysis of what physicalism "is", before canvassing various physicalist positions one might hold. I will outline the basic concept-space occupied by physicalist theory, including the commitments I alluded to earlier which any given physicalist theory must satisfy (while remaining agnostic, for now, on just how they might be satisfied). From there, section 2 will elaborate on reductive physicalism and NRP, canvassing motivations and trade-offs of said positions, at least those which are salient to this discussion[56]. Section 3 explains the argument, or *arguments*, from multiple realisability. I will note some problems or limitations of MR arguments. However, these are not of the putatively *fatal* sort; as already noted, this paper proceeds under the assumption that no fatal issues exist for at least some basic MR argument.

Section 4 considers those views which have emerged to deal with MR and outlines just how they do so; variously token identity, realisation in two forms, constitution and composition.

[56] Since a full accounting of every theoretical advantage and disadvantage of just, say, type-identity theory would require far more space than this paper can accommodate; my scope is restricted primarily to multiple realisability and causation for purposes of brevity and relevance.

Having laid this groundwork, section 5 offers an analysis of where these views seem to go wrong. Importantly, I seek to outline an underlying mistake endemic to all such approaches so far advanced. Having identified the cause of our NRP woes, we will then be positioned to treat them. This treatment is found in section 6, where I highlight some reasons in favour of expanding the NOAA constraint, and just how this allows for a less modest NRP. To do so in the first place, I propose a novel but straightforward division, which I map to two sorts of NRP, the standard view held by a variety of conventional non-reductive positions, and my proposed formulation.

As noted, this division is between "thin" and "thick" forms of non-identity, corresponding to conventional, "thin", NRP, and proposed "thick" NRP. In section 7, I begin with an analysis of Pereboom's proposed relation, concluding that it once again fails for reasons to do with thin non-identity. I then argue that moving to "thick" NRP is wholly *permissible* for the NRPist, that thick non-identity can fit within the intuitive picture of NRP. Finally, I offer two potential formulations of NRP which feature thickly non-identical properties and relations, demonstrating that these formulations also correspond to the NRP side of a technical division between NRP and naturalistic dualism, to assuage worries that I might be "sneaking in" dualism. Section 8, finally, offers some reasons for thinking that thick NRP is not only permissible, but desirable; a salve to the problems thus-far highlighted, and a source of other theoretical boons which I elaborate. I then conclude with some thoughts on where this modified view leaves us and provide suggestions for future work in this area.

3.2 What is Physicalism?

We have already discussed the basic physicalist picture, yet before moving to consider NRP, it will be worthwhile spending some more time on how physicalism is to be broadly understood. Explicating certain motivations and commitments of physicalism will allow me to demonstrate, later, how they might be significantly satisfied by what I referred to in the introduction as Thick-NRP.

As we know, physicalism does not pick out just one well-specified position, but it is better understood rather an umbrella under which a number of distinct positions fall. This nevertheless invites the question of just what physicalism "is"; what binds the set of physicalist positions together, and disqualifies others from the category. I have already offered various quotes "summing up" the basic physicalist view, and they generally resemble the following (what I will call the "First pass" view):

(1) **First pass**: Physicalism is the view that *everything is physical.*

As Sturgeon notes, 'Physicalism is the view that actuality is exhausted by physical reality' (Sturgeon, 2000, p.121) To offer a more elaborated example of something much *like* the First Pass: 'Physicalism is the view that the world is the way it is, because the physical world is the way it is. All the facts, including all the mental facts, are fixed by the physical facts' (Kallestrup, 2006, p.459). This "fixing" of the world *in toto* through "fixing" the physical, as seen in Kallestrup's definition, is a metaphysically non-committal formulation of what we can observe in other typical attempts to sum up the "basic" physicalist view. See also

Loewer; '[physicalism is] the thesis that all God had to do to create our world was to create its physical facts and laws; the rest followed from these' (2009, p.39)

Beyond this, Stoljar offers up the following formulation of the "standard" definition-in-brief:

(2) **The Starting Point View**: 'Every instantiated property is necessitated by some physical property' (Stoljar, 2010, p.5).

Stoljar goes further in explicitly detailing the idea of a 'physical property' for the Starting Point View thusly;

'F is a physical property [iff]:

A: F is one of the distinctive properties of physical objects; and

B: F is expressed by a predicate of physical theory; and

C: F is objective; and

D: F is a property we could come to know about through the methods of science; and

F: F is not one of the distinctive properties of souls, ectoplasm, ESP, etc' (2010, p.57).

How "physical" is best understood is not our focus here but we ought to keep this conception in mind, since it broadly captures the sort of intuitive understanding of physical properties held by most I will engage with in this paper, and myself[57].

Finally:

(3) **Shoemaker's Definition**: '[Physicalism is] the view that all states and properties of things, of whatever kind, are physical or physically realized' (2007, p.1)

To be explicit, all of the above are typical ways of constructing a definition of physicalism, and all attempt to outline a suitably close relationship (whether identity, or some other relation) between the apparently or obviously physical entities, and all instantiated higher-order or higher-level entities[58].

Nothing in the first pass, or selected quotations, is clearly philosophically *faulty*[59], and we are certainly capturing something important here, if only just the foundational intuition or "picture" of reality for those who identify as physicalists. Yet, they seem to underdetermine

[57] One can oppose part, or all, of this conception of physical properties. Indeed, I offer a critique of this common conception of the "physical" in another paper. Nevertheless, here I understand physical properties to be "objective", expressed by predicates of physical theory, and definitely not ectoplasm, souls, or so forth. Perhaps ectoplasm could be real, and could feature in some future physics as explanatory. I talk much in this paper about the "intuitive physicalist picture", and I think F for Stoljar here shows inclusion of such within the postulates of some future physics will *not do* to include them within the intuitive picture for physicalists. If our best science somehow includes causally efficacious souls or dualist-minds, that would be counter to physicalism, at least for many. And it is that intuitive picture with which this paper is concerned. Though, see Hart (1988) for philosophical development of interactionist dualism wherein "psychic energy" could causally interact with fundamentalia of current physics in modellable, lawful ways.

[58] Strictly, my "first pass" itself does not do the work of specifying such a relationship, but any attempt to develop this definition beyond the simplicity of "everything is physical" will invariably do this work.

[59] I leave aside worries about the definition of "physical" itself. Those are discussed at length in my earlier paper, on Causal Closure arguments. Nothing here rests upon this.

what most really want “physicalism” to capture. If nothing else, the first pass looks impoverished. While it might capture the *core* intuition, it *does not* capture commonly related commitments to be elaborated on in the next section, nor does it say anything about the good- and bad-making features of any physicalist position *qua* physicalism.

So, more development is needed here, in both elaborating and constraining “physicalism”. In particular what *must* a theory accept in order to be called Physicalism, or to hang onto its physicalist *bona-fides*?

For a start, we might informally say that the most important commitment of any physicalist theory must be to avoiding non-physical additions-to-being. Physicalism is *meant* to be ontologically cheap (if not “free”), such that the existence of the physical properties is sufficient for the existence of all other existent properties.

A salient ground of, or motivation for, physicalist theory is causal closure, or the completeness of physics, related principles which find their way into a host of arguments for physicalism, or *contra* non-physicalism. Thus, the primacy, if not absolute-supremacy, of physical causes, ought to feature in any consideration of central tenets or constraints upon physicalist theory. Such grounds, and commitments, are explored further in the next section.

3.3 Delineating Physicalist Theory

I will now outline what further commitments must necessarily be held by a position in order for it to "qualify" as physicalist, as I see it.

I take it that for any theory to be rightly called physicalist, it must satisfy these commitments. However, so satisfied, we may also talk about ways in which certain theories better satisfy these commitments, or better fit the intuitive picture generated by the prior theses and these attendant commitments.

These two commitments are outlined as follows:

(C1) **Metaphysical**: For any higher-level property, higher-level property is Nothing Over And Above the Physical[60].

(C2) **Causal**: Every physical event which has a cause, has a sufficient physical cause.

(C2) seems relatively uncontroversial[61], and I do not think much needs to be said for the causal constraint on physicalist theories. Some may reject this formulation on the basis of their metaphysics of causation, or events, but I bracket these considerations since offering a

[60] I state "Higher-level Property" instead of "the mental" or "mental event" since the physicalist is committed not just to the mental properties being nothing over and above the physical properties, but to all higher-level properties from biology to economics being likewise nothing over and above.

[61] I put aside concerns about consciousness and quantum collapse here- some physicalists may think one can provide a story explaining the role of consciousness reductively even if consciousness plays some role in quantum events, while others may take the existence of such a phenomena to undermine physicalism. Nothing here is germane to the ongoing discussion, however.

full accounting of every way in which (C2) could be formulated would require its own work. In crude, perhaps more metaphysically neutral language, (C2) is meant to convey that for any concrete physical thing, or entity, or property, its diachronic origin will also be physical.

(C2) is a canonical, classical grounds for physicalism (Papineau, 2001, Kim, 2005), and denying it is tantamount to denying physicalism[62], though one can certainly hold (C2) without being a physicalist. If there were sufficient non-physical causes of physical effects where there were no corresponding sufficient physical cause, this would entail causally efficacious non-physical entities or properties, which would undermine the First Pass, and any other reasonable definition of physicalism. NRP, in some forms, might permit of mental causes of physical events, but these causes are, in some sense, derivative, of other physical events or property-instantiations. This seems a very reasonable caveat, but it *is* an important one, which I will consider later. We can now move to consideration of (C1), or NOAA.

'What does 'nothing over and above' mean? This slippery phrase has had a lot of employment in philosophy, but what it means is never explained by its employers' (Van Inwagen 1994, p.210)

NOAA talk is ubiquitous in physicalist discourse, from statements in myriad publications that some physicalist position ensures the mental properties, objects or events, are *nothing over*

[62] Most formulations of causal closure are not sufficient for physicalism (Lowe, 2000, p.573-576), though they would appear to be necessary for it; at least, so long as there is physical causation in the first place!

and above the physical properties[63] (Kim, 1992, p.6, McLaughlin, 2007, p.150, Menzies, 2013, p,59), to concerns that certain relations or theories might *fail to ensure* the mental properties are nothing over and above the physical (Noordhof, 2013, p.112). Indeed, standard NOAA is sometimes, not unjustifiably, taken to be *the* mark of any physicalist theory. It is easy to see how Dependence relates to NOAA. As a reminder:

[Dependence]: All instantiated properties, including mental properties, are entirely dependent upon instantiated physical properties

Dependence aims to capture the ontological innocence of the NOAA commitment, noting that any sort of instantiated property is necessarily dependent upon some physical property, not beyond or distinct from the base physical properties.

One might suppose that (C1) conceptually follows from the definitions offered in the prior section since, if the physical properties "fix" *all* of the properties, then all properties might be thought of as determined by, and dependent on, physical properties. At least, this seems a natural way of understanding "fixing" talk. Necessitation and realisation talk, as found in Stoljar and Shoemaker respectively, make this *absolute* dependence more explicit, offering an 'ontological picture' of 'derivative and dependent' mental properties (Kim, 1992, p.6). This constraint warrants closer examination, however. We ought to begin by considering *exactly* what is meant by NOAA, before moving to those reasons and intuitions which motivate and *delineate* the constraint.

[63] Armstrong coins the term 'ontological free lunch' (Armstrong, 1997, p.12), for any entity which is not over and above pre-stipulated facts or properties.

As Pereboom states, 'the truth of physicalism [entails] there must be a sense in which the token causal power of mental state M would be "nothing over and above" the token causal power of microphysical state P' (Pereboom, 2011, p.135). So, one element of what it means for x to be nothing over and above y, is for x to have no causal powers, or unique causal power, besides those possessed by (and "inherited from" or shared with) y. This seems to be only part of the puzzle, however, since NOAA is not just a position on causation. If "standard" NOAA is true, then there are *no* higher-level properties which are distinct from their microphysical bases, in any "meaty" sense. There are no fundamental non-physical properties, no fundamental instantiated properties at least. In other words, NOAA guarantees aforementioned ontological *innocence* (Lewis, 1991, p.81); that there are no additions to being, excessive metaphysical import, or commitments to entities *beyond* the basic "lower-level" entities.

> 'The fusion is nothing over and above the cats that compose it. It just is them. They just are it. Take them together or take them separately, the cats are the same portion of Reality either way. Commit yourself to their existence all together or one at a time' (Lewis, 1991, p.81)

Lewis notes that if one is already committed to the existence of cats, a commitment to cat-fusions (groups of cats), is no real additional commitment (p.81-82). Likewise, for NRP, a commitment to the existence of certain microphysical properties (and to NOAA itself, of course) requires no additional commitment to "higher-level realised or composed or [some

other suitable relation]" properties[64]. This, at least, is the aim of NRP, howsoever it is formulated.

That typically construed NOAA entails no additions-to-being, has some apparent conceptual fallout. Emergent properties are ruled out entirely, when understood as properties distinct from their microphysical emergence-bases, and usually with distinct, novel, causal powers. Here is Melnyk on an objection to constitution-based NRP;

'this sufficiency might be sufficiency **in accordance with a fundamental law of physical-to-mental emergence** whereby, if an event of p's physical type occurs, then an event of m's mental type occurs; and if it is, then (mental) m's being constituted by (physical) **p won't entail that m is nothing over and above the physical**' (Melnyk, 2008, p.1295, emphasis mine).

I will say more about emergence and its relation to physicalist theory later. For now, I only wish to make clear the strong *prima-facie* tension between emergent properties, and NOAA.

Why is NOAA *desirable* as a constraint? I take there to be a few interrelated motivations for its widespread popularity in physicalist theory.

[64] There may be non-causal, non-phenomenal, properties, held by some realised property and not its realisers, or vice-versa. For instance "the property of being realised by [microphysical state]", or "membership of set [y]" where [y] is some set which excludes realiser or the realised. We can also consider the possibility of extrinsic properties, or non-qualitative properties (Langton and Lewis, 1998), which *might* similarly differ between base-physical properties and mental properties. Whether these are serious, thoroughgoing properties, or undue reifications is up-for-grabs in debates between metaphysicians quite generally, but no conclusion here is likely to be truly problematic for the NRPist. Such conceptual possibilities are noted only for the sake of completeness. Whether or not these entities "come for free" in our metaphysical theories, this is a level removed from our current discussion. Put simply, nothing rides on the placement of such properties in our ontology, for purposes of NOAA or this paper.

NOAA is a clear part of the "intuitive" physicalist picture, and a given for physicalist worldviews. Anyone would simply fail to be a physicalist if they thought, so it goes, that mental properties were in fact over or above physical properties. As I see it, NOAA is not best understood as a technical or metaphysical notion[65], but something that we bottom out at and a proposition that our theories need to do justice *to*.

Besides that, an obvious note is its relation to the causal constraint, (C2). Causal-closure principles are developed in several ways, and all precise formulations have been substantially critiqued. Yet, there is a core insight that the physical properties are sufficient for bringing about any and all later physical effects. With this as a given, it is easy to adopt an intuitive view that the physical is doing all the work, and anything besides must be either inconsequential or non-existent. Rather, nothing over and above the physical is causally or explanatorily *necessary*. While this does not suffice as, say, a deductive argument *for* NOAA, it might function as a partial explanation for adoption of the constraint.

Further, it represents a commitment to ontological innocence. Indeed, I might even have framed (C1) in terms of ontological innocence, but for the more standard, frequent, use of NOAA across work in the field. Ontological innocence is also often portrayed as a more general commitment for philosophers, or a good-making property of theories, beyond physicalism (Lewis, 1991, p.68, 81-82). Innocence, or innocent theory, does not '[add] to the list of entities whose existence you are committed to' (Hawley, 2014, p.3). Thus, the *more*

[65] Indeed, I have not tried to characterise it as such. As I note, I take it to be "part of the intuitive picture" for physicalism; this may mean it looks "slippery" to Van Inwagen, but it will have to suffice that we can explain what it entails, and prohibits, rather than cashing it out into further primitive notions.

innocent a theory, the *fewer* entities it commits one to[66]. The attraction to ontological innocence might be fundamental for some, or otherwise rooted in the application of Occam's Razor which pushes, all things being equal, toward the simplest theory with the fewest ontological commitments (Quine, 1981, p.10, 144, Baker, 2016). I do not intend to argue against the merit of ontological innocence here; I will however note, as others have (Quine, 1984, p.13, Lowe, 2009, 208-217), that innocence is one positive feature a theory might hold, to be balanced against considerations such as explanatory strength, coherence, and whatever additional benefits might attach to particular commitments.

Having established what this commitment looks like quite generally, we can consider how it is realised in relation to physicalist theory. The physical (typically, the microphysical) entities are those to which we are all already committed. We can recall the cat-fusion case from Lewis; any good physicalist ontology does not wish for a fundamental commitment to cats, but to the microphysical entities, whatever they turn out to be[67]. It *avoids* commitments to other, non-physical entities, like immaterial mental "stuff"[68]. Consequently, standard-NOAA is best-satisfied by either identity, or some relation likewise "innocent"; it does not threaten to import any sort of dualism "via the back door" by permitting the postulation of novel non-physical properties. The mental properties, in an ontologically innocent physicalist

[66] Though often taken for granted, it bears stating that it is qualitative, not 'quantitative' (Lewis, 1973, p.87) parsimony at issue. It is not the mere number of entities that we are interested in, but the postulation of sorts or varieties of entity. To give a simple example, a physical theory which postulated only Monads would be more parsimonious than the theory postulating Monads, Anti-Monads, and God Particles, even if the former theory posited ten times more monads, or particles generally, than the latter.

[67] Classically, this was *matter*, but physicalism has grown out of 'materialism' (Loewer, 2001, p.37) as it has continued to respect and thereby update its ontology alongside, our best theories in physics.

[68] I sometimes use "stuff" as an informal, metaphysically neutral way to refer to certain classes of entity, when nothing rides on the precise metaphysical cashing-out of a given claim beyond this.

theory, are meant to be a “free lunch”, as the saying goes, no matter what relation between physical and mental satisfies the constraint.

A related motivation for adopting standard-NOAA emerges out of this; standard-NOAA does not permit a build-up of “unwieldy” properties at various levels of reality. The thought is that the physicalist can talk intelligibly about mental properties, amongst other higher-level properties, without thereby committing themselves to the “fundamental” existence of said entities. At the “bottom level”, all that *really exists* is the physical *stuff*, on an intuitive physicalist understanding. Borrowing the terminology of Hodes (1990) and French (2016), we might talk of ‘thin’ and ‘thick’ commitments, with physicalists being thinly committed to mental properties, since the mental properties are talked about within the sentences of physicalist theory, while the *truth-makers* of statements regarding mentality, under physicalism, are invariably microphysical properties and their arrangements. Explicitly metaphysical worries about “unwieldiness” of higher-level properties aside, there is also a commonplace belief that higher-level forces and processes will invariably be subsumed into the small set of fundamental, physical forces, as has *historically* been the case (Papineau, 2001). This functions as a broad motivation for physicalism, and as a *particular* motivation for NOAA.

So far, we have considered how the causal constraint gives *cause* to suppose that non-physical properties are not anything causally over and above the microphysical properties, that there ought to be no non-physical causal influence “downwardly” upon the physical. Further, we have seen the motivation vis-à-vis ontological innocence. Namely, if non-physical properties must be posited or discussed, they can at least incur no additional

commitments, whether through being identified with the base-physical properties, or related in some similarly non-committal, non-weighty manner. These are not just motivations for NOAA, either; they help us to understand what standard-NOAA is, beyond any handwaving about "Over and Aboveness" itself, or metaphors about free lunches. Namely; no downward causal influence, especially influence that cannot be modelled as flowing directly from base-physical causes, and no, or *absolutely minimal*, positing of non-derivative mental (or otherwise higher-level) properties in physicalist ontology.

I do suspect that there is one other motivation for standard-NOAA[69], though it is not one that I think is especially useful for illustrating how NOAA functions as a constraint, nor for delineating its bounds, as the causal and innocence motivations are. Rather, it *underdetermines* NOAA[70], though I think it might even be conceptually necessary to hold, if one is to hold NOAA. It is also, unlike the others, not readily referenced in the literature, but this might be because it is simply so *obvious* that the need to note it is not generally present. Alongside Nothing-Over-And-Aboveness, I call this the Nothing-Left assumption, summed up as follows:

"Take the physical away, and there will be nothing left".

[69] And revised-NOAA, as will be demonstrated in a later section.
[70] It might not even be sufficient for physicalism, though one could not be a physicalist without it.

In other words, wherever there is physical and mental *stuff*, removing the physical stuff, whether by decomposition or annihilation, also removes the mental stuff[71]. Where the relevant neuro-physical property, for instance, there will be no pain property.

If this seems vague, that is because it is; I am only trying to capture a foundational intuition here, one which seems to lie behind NOAA. Indeed, it seems reasonable to treat Nothing-Left, properly understood at least, as a fundamental *component* of NOAA. It captures the primacy of the microphysical domain[72], both causal and ontological, while remaining neutral on the sorts of relation that might obtain between the physical and higher levels. It also does justice to a common-language reading of NOAA. For one entity to cease to be when another does, as a matter of necessity, indicates that the former "gets" its existence *derivatively* from the latter, at least when we restrict our scope to instantiated concrete properties.

Kim, for instance, explicitly links the 'ontological picture of mental properties as derivative and dependent' with 'suggestion that when we look at concrete reality there is nothing over and beyond instantiations of physical properties' (Kim, 1992, p.6). While Nothing-Left is not *usually* declared itself, the derivative role it frames mental properties within is found in standard formulations of physicalism:

[71] This could be read as a statement about the totality of entities; when all physical entities and properties in the universe, or the physical universe as a whole, are annihilated, there will be no mental entities or properties remaining either. I intend something more fine-grained; that specific physical properties being removed or altered, in whatever way, will also remove dependent mental properties. Some philosophers take it for granted that mental events can be temporally extended, across a number of physical dependence-bases at different times. I do not intend to rule this out here, but only wish to make clear that in the absence of any sufficient physical dependence-base, there will be no mental property.

[72] Melnyk provides this instructive point, which I take to capture how essential the primacy of the physical is to the physicalist picture: 'to qualify as a version of physicalism, non-reductive physicalism must honor the intuition that fundamental physics is the basic science' (Melnyk, 1995, p.371)

‘[All] God had to do to create our world was to *create its physical facts and laws*; the rest followed from these’ (Loewer, 2001, p38-39, emphasis mine)

That is, ‘the rest’ (p.39), like mental facts or properties can only appear on the scene in virtue of the “right” physical elements already being present. See also:

‘[All] facts obtain in virtue of the distribution of the fundamental entities and properties’ (Loewer, 2001, p.37).

Mental facts “get to exist” derivatively, in virtue of fundamental physical facts, mental properties in virtue of physical properties, and so forth: ‘non-reductive physicalism must **honor the intuition that fundamental physics is the basic science**, so that all other sciences, including their laws, must be dependent on, and hence derivable somehow from, fundamental physics’ (Melnyk, 1995, p.371, emphasis mine). To quote Sturgeon, contra the token dualist, the physicalist holds that mental properties are *not* ‘radically distinct’ (2000, p.5).

This should not be controversial to any physicalist; if mental properties existed independently, or could persist despite the destruction of related base-physical properties, this would be bad news for physicalism. ‘Indiscernibility with respect to physical characteristics [guaranteeing] indiscernibility in respect of mental characteristics’ (Kim, 2012, p.170-171) in other words, is what is required, or some formulation of dependence or *supervenience* between physical and mental properties.

I note, parenthetically, that Nothing-Left in isolation from standard-NOAA, would seem to permit varieties of emergence. I do not think some accommodation of emergence is as worrisome for a committed-physicalist as has been sometimes asserted (Kim, 1992, p.18), though more will be said about this later. Of course, NOAA does not state "nothing *but*"; only nothing over, and nothing above. Clearly, something other than the physical, or microphysical, is accommodatable or NRPists would have cause for concern about the mental elements of their theories, even as they now stand. While I have introduced Nothing-Left as both a component, and partial intuitive grounds, of NOAA, I intend to argue that we ought to loosen our NOAA commitment on NRP, to Nothing-Left itself. Given that NOAA is itself an intuitive notion, we might think of this as merely "raising the bar" on what theses we take to entail nothing-over-and-aboveness, although accepting it as a distinct and more permissive commitment would change little here. I discuss Nothing-Left and its merits later in this paper, but one benefit of endorsing Nothing-Left rather than standard-NOAA is worth mentioning immediately. Nothing-Left, while still an attempt at capturing an intuition, is more legible than NOAA. We can readily generate, for instance, necessary conditions such that a theory would satisfy Nothing-Left.

Necessary Condition: For any theory, if some instantiated mental particular could exist in a world without some physical particular as its dependence-base on that theory, Nothing-Left is not satisfied.

It is not clear to me, at least, that one can do the same for NOAA without simply recapitulating NOAA itself. By my lights, it is much more clear whether Nothing-Left is

satisfied by any particular theory, than NOAA. While nothing here is intended to undermine or dismiss NOAA discourse in metaphysics quite generally, we have a choice between two commitments of NRP specifically. Both seem to satisfy previously listed motivations, but Nothing-Left is clearer and more prima-facie explicable. It also opens the door to enormously beneficial theoretical work which NOAA, currently placed as the second commitment, (C2) of NRP, does not, as I will elaborate in later sections. For now, I leave (C1) and (C2), and consider whether there might be some further commitment of NRP along modal lines, before moving to discussion of present formulations of NRP and their satisfaction of these commitments.

3.3.1 Is there a Modality Commitment for Physicalism?

It would be fair to say the jury is still out on the modal status of physicalism. Some argue that physicalism must be a necessary truth, if it is to be true at all (Davidson, 1970). Still, others have advanced arguments for physicalism being better off as 'modally modest', leaving open the possibility that things could be otherwise, but denying that this entails dualism in the actual world (Sturgeon, 2000, p.154). Though I will have comments on modality and physicalism in relation to my later outline of "thin" and "thick" non-identity terminology, I do not take it that the modal status of physicalism, one way or the other, is as *clearly* a commitment as NOAA, causal closure, or naturalism.

Regarding the thin/thick distinction I refer to, I will note that NRP is amenable to being *modelled*, modally. In section 6, I explain how this understanding allows us one way of

distinguishing NRP in any form from naturalistic dualism, and how my "thick" model fits into the NRP side of this modal model.

3.4 Reduction, Non-Reduction, Type-Identity, Token-Identity

With modality considerations aside, we are left with (C1) and (C2). Type-identity theory (and reductionism), satisfies (C1) and (C2), quite uncomplicatedly, as discussed. It is naturalistic, committed only to physical causes of physical effects, clearly satisfies NOAA, and avoids any sort of radical distinctiveness through identity between higher-level mental properties and basal physical properties. However, type-identity faces a MR problem, to be explained in Section 4. Before that, Section 3 outlines type-identity as well as the positions physicalists pivot toward in the face of MR concerns.

There are various technical notions of reductionism "up for grabs", but our concern here is with *metaphysical,* or ontological reductionism, rather than explanatory or theory-reduction (notably, the historical case for multiple realisability (henceforth MR) was aimed at problematising theory-reduction (Fodor, 1974)). I have already glossed over reductive physicalism and type-identity, but here I offer some clear definitions, for clarity's sake and to delineate the positions precisely.

Ontological Reductive Physicalism: For every mental property, there is some physical property with which it can be identified (Kim, 2012, p.174, Searle, 2004, p.701)[73].

[73] In other words, the properties picked out by the predicates of psychology/[higher-level theory] are identical to certain properties picked out by the predicates of our physical theory We need take no stance on whether those predicates themselves can be "translated" to predicates of the more fundamental theory. Debates about bridge-laws and the logical derivability of theories are set to one side in this discussion.

Type-Identity Theory: The reductive identification of mental properties with physical properties (Kim, 2007). For every mental property, there is some physical property to which it is identical, or in Fodor's terminology '[type physicalism is] the doctrine, roughly, that every property mentioned in the laws of any science is a physical property' (Fodor, 1974, p.100).

Ontological reductionism and type-identity rise and fall together; if ontological reductionism is false and one or more mental properties are not identical to some physical property, then type-identity theory must also be false since it states that every mental property is identical to some physical property. I spell them out separately only because they are distinguished in the literature, and their respective converse positions do come apart in a notable way. It is possible to identify as a non-reductive physicalist *without* being a token-identity theorist. With one aim of this paper being to carve out a NRP position which does not involve identity, it is worth getting clear on how NRP in the literature involves no reference to identity.

The Fodorian line neatly sets up our definition of *token*-identity:

Token-Identity Theory: 'Token physicalism is simply the claim that all the events that the [special] sciences talk about are physical events. (Fodor, 1974, p.100).

I understand events here, and throughout, in terms of property-instantiation-at-a-time (Kim, 1993, p.37), though the question of appropriate metaphysical framework for events is

contentious. At this point, however, little depends on what view of events you are inclined to adopt, so long as they can be bearers-of-properties such that we can talk intelligibly about the identification, or not, of instantiated mental and physical properties.

Non-Reductive Physicalism: Mental/Higher level properties are not reducible to the properties of fundamental physical theory.

The NRPist remains committed to (C1) and (C2), just as any physicalist. They also have additional challenges to grapple with, namely ensuring the causal efficacy of the mental, in light of the Exclusion Argument[74], and avoiding recurrent MR worries however they formulate the relation between mental and physical properties.

With that in mind, section 4 examines the standard multiple realisability argument, which NRP seeks to avoid and which, as I will contend, "thin" NRP fails to adequately handle.

3.5 The "Master" Multiple Realisability Argument

The intuitive notion behind multiple realisability is that certain mental states could occur despite a difference in physical substrate. MR is typically motivated by either empirical appeals to multiply realised mental states inter- or intra-species, or appeals to the theoretical possibility of distinct realisations of a given higher-level property.

[74] I offer a model for NRP dealing with this causal dilemma in another paper.

I could, for instance, have a pain state, or some other psychological state, identical to that experienced by some silicon-brained alien, or artificial intelligence, or some other entity with an entirely distinct neural architecture, an 'automat[on]' (Fodor, 1974, p.105) perhaps. It is, elsewhere, conjectured that different people could experience the same pain states despite variance in the parts of the human brain which realise said states, in light of findings relating to neural plasticity (Shapiro, 2002). Or, that certain animals could undergo mental states indistinguishable from ours (Block and Fodor, 1972, p.238). I do not take any stance in this paper on the truth-value of these propositions; I only note that they are some common grounds from which general MR claims are derived. If a given mental kind, or property-type, or state, is common across two or more undergirding physical states, then it is not reducible to any one of them[75].

The reductive physicalist supposes an identity between mental and physical properties. Following from this, as Kim notes; 'If a higher-level special-science property M is identical with a physical property P, M and P must be coextensive; that is, the correlation "M occurs if P occurs" must hold' (Kim, 2012, p.174). We then observe that 'there is no single physical property P such that M occurs if P occurs. Rather there is a multiplicity of physical properties P1 , P2 , … such that M occurs if P1 occurs, M occurs if P2 occurs, and so on' (p.174). If mental properties are multiply realisable, then there is not identity between mental and physical properties.

[75] I assume throughout that identity is not contingent, though it should be acknowledged that this assumption is neither universal nor uncontroversial (Schwarz, 2013). Serious argument for *not* endorsing contingent identity would require its own paper, and those who do endorse it may have quite different responses to multiple realisability than anything discussed in what follows. I put this aside, therefore, and address all that follows to physicalists who take identity to be necessary.

Kim is simply asserting that identity entails co-extension and we *seem to observe* in these cases that co-extension fails to hold, meaning that there cannot be identity.

We can fashion a very general "master" argument, which avoids reference to the mental per se, to illustrate the shape of MR arguments. Following this, I will "plug in" the mental-physical relation, which the intuition itself is targeted at. In the following premises and conclusion, x and y refer to events or instantiated properties.

P1: Multiple Realisability Claim: *Some* higher-level x is multiply realisable by lower-level, or *fundamental* y.

"Realise" in P1, and P2, does not refer to any philosophical term of art, but only to the idea that some property can be co-extensive with a number of lower-level properties. No commitment to any particular view of the relation between these properties is intended, as yet.

P2: Strict Identity: For x=y to be true, x must not be realisable where ¬y. That is, if x is multiply realisable, it cannot be identical with any one y.

If x is strictly identical to y, then x will not exist in any situation or possible world where y does not exist. However, if x is instantiated when *either* property y or some distinct property z are instantiated, then x is not identical with either. Where P1 holds, this will be the case.

C: Failure of Identity: *Some* higher-level x are not identical with lower-level, or fundamental y.

We can reconstruct this with regards to mental and physical properties as follows:

P1: Some mental properties are multiply realisable by distinct physical properties.
P2: If some mental property is identical to some physical property, the mental property must not be realisable by any *other* physical property.
C: Some mental properties are not identical with physical properties.

Briefly, if mental properties are not identical with physical properties, then mental properties are also irreducible to physical properties. In particular, there is no identity to be found between mental and physical kinds or types. Some alternative conception of the relation between mental and physical is required:

> 'The main point of MR that is relevant to the antireductionist argument it has generated is just this: mental properties do not have nomically coextensive physical properties, when the latter are appropriately individuated. It may be that properties that are candidates for reduction must be thought of as being realized, or implemented, by properties in the prospective reduction base' (Kim, 1992, p.6)

3.6 From Type to Token Identity: The Positions

If MR concerns are justified, as this paper assumes, we need to get away from type-identity! However, if we are deprived of the identity relation, some account of how mental and physical relate is required for any NRP theory. It is not sufficient to stipulate the truth of physicalism conjoined with irreducibility. We need an account that incorporates irreducibility, while remaining physicalist; that is, satisfying commitments C1-C2.

Reductive physicalism achieves this straightforwardly (absent irreducibility, of course), as noted. If Physicalism is, fundamentally, the claim that all entities are, "ultimately", physical, then type-identity reductive physicalism achieves this by postulating identities between higher-level properties and the properties of microphysics[76], or the properties picked out at some higher macro level of physics. Everything is "ultimately" physical, on this model, because every higher-level property/fact/event is identical with some microphysical property/fact/event. It also comes pre-equipped with a satisfying answer to causal-exclusion worries! It is entirely clear how mental properties get their causal "oomph" on the reductionist view; by their identification with causally efficacious physical properties.

So where do we go, if committed to irreducibility *and* C1-C2? This section will proceed to canvas a range of NRP formulations which, I take it, are broadly well-regarded, canonical responses to MR. I will then spell out what distinguishes these views and NRP formulations generally, from naturalistic dualism. In the last section, I outlined the shape of multiple

[76] Here I purposefully elide a host of canonical worries with definitions of "the physical" as a load-bearing part of physicalist theory, largely because they're orthogonal to the current discussion. Somewhat related concerns, however, will play a role in a later section of this paper.

realisability problems, which I will argue re-assert themselves for these NRP formulations. The canvas of formulations which follows will provide us with a range of current physicalist responses to MR, and I intend to demonstrate what these views have in common and how the outlined sort of MR problem they attempt to respond to reasserts itself for *them* in following sections, precisely because of a feature they share- thin non-identity. I will also introduce naturalistic dualism and distinguish it from NRP within this section, showing why these formulations do not constitute naturalistic dualism, and providing us with the resources to later distinguish my own proposed solutions from naturalistic dualism. Without belabouring a concern I intend to address later, this will be important since my suggested formulations hinge on loosening a commitment of NRP which might reasonably be thought to delineate NRP from naturalistic dualism. Starting here, I hope to show that this loosening should not concern physicalists who do not wish to accept any sort of dualist position; my proposals will be shown to be *at least as distinct* from naturalistic dualism in significant respects as the NRP formulations canvassed in what follows.

First, a fairly typical conjunction of views, in light of these considerations, is type-dualism, or a property-dualism (where every mental property F is distinct from every physical property F* (Yablo, 1992, p.269-270)) with *token*-identity, and supervenience. All mental property *tokens* are taken to be identical with 'physical property tokens' (Schneider, 2013, p.136), such that there is an identity between mental and physical *events*. NRP often, though not necessarily, couples this with a causal-role, functionalist or computationalist account of mental properties (p.136). Realization accounts are essentially *functional* in nature:

“Standard” Realization: Property X realises Property Y just in case X fills the causal role occupied by Y (McLaughlin, 2007, p.151).

Shoemaker offers a distinct account of realisation, which I outline here for later reference:

Subset-Realization: ‘Property P has property Q as a realizer if the forward-looking causal features of P—its aptness to contribute when instantiated to the production of certain effects—are a subset of the forward-looking causal features of Q [and] the backward-looking causal features of the realizer [are] a subset of the backward-looking causal features of the realizer properties’ (Shoemaker, 2013, p.41).

Otherwise, some NRPists move to composition or constitution, whereby the mental properties are constituted (or composed, see Charles (1992) and Pettit (1993)), by some physical properties in a token instantiation (Crane, 1995, p.9). For instance:

Causal-Powers/Property Constitution: 'The causal powers of a token of kind F are constituted of the causal powers of a token of kind G just in case the token of a kind F has the causal powers it does in virtue of its being constituted of a token of kind G' (Pereboom and Kornblith, 2004, p.714).

So we have, broadly speaking, two camps of account; realisation (including functionalist and computationalist theories), and constitution (or composition), in addition to “bare” supervenience with token-identities, or the identity of mental and physical causal powers. All these views aim to dodge MR worries while ensuring standard-NOAA.

Before moving on to a critique of NRP, or "thin" NRP, as I describe it, I wish to make clear one thing which NRP (in all the above forms) is *not*, or at least, which it is typically distinguished from- namely naturalistic dualism. Several points of divergence exist:

(1) Functionalisation of mental properties

The NRPist usually accepts some sort of causal role for mental properties (Schneider, 2013, p.136), whereas naturalistic dualists take qualia to be non-capturable by any sort of functional or causal analysis (ibid). This acceptance of epiphenomenalism is elaborated in point (2).

(2) Qualia as 'fundamental' (Chalmers, 2004, p.630) or 'irreducible features of the universe' (Schneider, 2013, p.136) for naturalistic dualists.

In a sense, both NRP and dualists are committed to irreducibility, thus the language of "fundamentality" perhaps better captures the relevant dividing line. Qualia are real, "serious" additions to being, part of the furniture of the universe as the metaphor goes, in need of "adding in" even once the physical facts are all fixed. In terms of motivation for this, naturalistic dualists come apart from the conventional NRPist also; the dualist is motivated by arguments such as the explanatory gap, appeals to p-zombies, and the knowledge argument. This can lead to the conclusion that such qualitative phenomena are also *epiphenomenal* (Jackson, 1982, p.134), removable from the causal (and physical) picture of reality without fuss, as noted in (1). This epiphenomenal dualism

does not always extend from the phenomenal to the *psychological,* which generates what could be called the "phenomenal/psychological divide", to be addressed in more detail later.

(3) The modal strength of the supervenience relation between mental and physical properties; for NRP, this is usually metaphysical strong supervenience, while for naturalistic dualism, it is merely nomological strong supervenience (Baysan, 2020). For NRP, 'a given supervenience base property metaphysically necessitates the supervening mental property' (ibid), for naturalistic dualism 'it only nomologically necessitates the supervening mental property' (ibid).

(3) will be returned to later, as I will demonstrate that my thick-NRP maintains parity with the modal status of standard, or thin-NRP. Distinguishing NRP this way, particularly my novel positions, from naturalistic dualism is important as previously noted since one might worry, reasonably, that "thick" non-identity and Thick-NRP, may begin to "look" far closer to dualism than thin-NRP.

My next two sections constitute an analysis of where thin-NRP goes awry. First, I will consider how thin formulations, which I contend principally go wrong in attempting to cleave close to (C1)/standard-NOAA through thin non-identity, fall prey to the same sorts of MR concern which motivated their adoption.

As one note before getting to this, I do not intend to discuss supervenience-physicalism itself. This is simply because supervenience is typically conjoined with some sort of

realisation relation, or with token identity. Supervenience alone is generally taken to substantially underdetermine physicalism[77], only '[tracking] certain patterns of modal covariation without accounting for why these patterns hold' (Stenwall, 2021, p.11778); it does not tell us why these mental-physical relations exist, or what they really are. Given that our concern is with presenting a formulation amenable to NRPists, I think this is sufficient to forego discussion of "mere" supervenience-physicalism.

3.7 MR's Revenge, Tokens and Causal Powers

We can begin with token identity itself, which satisfies (C1) via the identification of mental and physical events[78], while maintaining property type-dualism as noted. After this I move to see how MR-related problems impact the other discrete sorts of NRP canvassed in the prior section. While noting that problems arising from MR-acceptance are central, I also consider other evident issues with these views, especially here for token-identity, as follows.

If one adopts a Kimean property-exemplification account of events, then one runs into immediate trouble with token-identity:

"Take any token psychological event under the property exemplification view. It will be an object having a psychological property, P, at time t. If this is also a token neurological event,

[77] I do not discuss this in detail for purposes of brevity, but see Stenwall (2021), Crane (2001), and Kim (2012, p.172); for why supervenience is insufficient for 'minimal physicalism' (ibid).
[78] There is room for dissent on this point. Kim argues token physicalism may in fact fail to respect the primacy of the physical in a way supervenience and other NRP formulations do not; it simply asserts identity, without offering anything to demonstrate the 'primacy' (Kim, 2012, p.179) of the physical within what I have called the physicalist intuitive picture. Mental tokens are identical with certain physical tokens, not necessarily derivative and dependent.

there must be a neural property, N, such that the same object has N at t , and the events, the object's having P at t and its having N at t , are one and the same event. From the identity criterion of events as property exemplifications, it then follows that psychological property P = neural property N! On this view of events, then, token identities imply type identities, and token physicalism will collapse to type physicalism" (Kim, 2012, p.175)

Put simply, the psychological-neural event in question instantiates a psychological property and a neural property. It further stipulates these properties must be *identical* to one another, as only one property is instantiated in-the-object-at-a-time. Thus, on a property-exemplification view of events which is, at least, plausible (and a seemingly good model for individuating events in a suitably fine-grained manner), token identity ends up '[implying] type identities' (p.175)[79].

To illuminate this worry, and how it interacts with MR, type-identity was problematised precisely *in virtue of* the apparent non-identity of psychological and neural property-types across a host of cases. Yet the psychological property instantiated (or "exemplified") in these token events (on the property-exemplification view) can be multiply realised across just such a host of cases, despite being putatively identical with the neural property in question. Thus, any MR issue for type-identity also holds for "token-identity" on this account of events.

[79] Horgan (1984, p.322-38) makes a similar point here

Let us take it, *arguendo*, that token-identity does not simply collapse into type-identity, as many believe. For instance, consider the account of property-instantiation proffered by Block and Fodor (1972)[80]:

'[A] token of type x may be identical with a token of type y even though x and y are distinct types.'

An MR issue would appear to remain, in the following formulation:

P1: For some token M, realised by some neural token P, M could have been realised by some alternative neural token, P*.

P2: Any token [x] can only be identical with itself.

P3: P and P* are not identical, neither does P realise P* nor P* P.

P4: If M can be realised by P*, M is not identical to P.

C: M is not identical with P

To the extent that a given mental event could be identical to one realised by distinct neural underpinnings, the argument is sound; if M=M and P=/=P* then M is not identical to P, or P*.

Certainly, there are objections against this line of argument. If such objections are premised upon a straightforward denial of the possibility that two mental tokens could be identical

[80] Or, perhaps this pitfall can be avoided via Shoemaker-realization, to be discussed shortly.

while instantiated by distinct neural tokens *contra* P1, regardless of phenomenology or appearances, it is not clear that this would not run equally well for the standard MR argument. That standard argument, however, possesses intuitive force which token-identity physicalists accept. If one is motivated to *accept that*, then *prima-facie* one ought to accept this, barring very good reasons for these cases coming apart. This goes for various arguments against this construal of MR as applied to tokens or events. If we are willing to accept MR for type-physicalism, then we need clear reasons for denying its extension here. As it stands, the mere assertion of brute identity appears *ad-hoc*; mere correlation or covariation will not do.

A perhaps more promising response may be to deny that event-identity is ever so easily procured[81]. M-realised-by-P and M-realised-by-P* might be *similar*, but perhaps it should be regarded as controversial to suggest that these distinct mental events, with distinct neural underpinnings, are the same. Nevertheless, I think our intuitions ought *minimally* to be weighted in the other direction. This objection, regardless, would not ground the identity of the token mental and physical events, and would only offer a route for the token-identity theorist out from accepting their non-identity. P1 seems extremely plausible, certainly if one is already inclined to accept MR reasoning in general; it seems our sympathies should be against the token-physicalist position.

Further, Pereboom (2002, p.132) suggests a temporal revision of this argument, wherein a mental event or state persists across time despite a change in its neural realisers. Suppose a

[81] Thanks to Dr William Sharp for suggesting this during a seminar.

persisting belief obtains across time (p.134), for instance, or even a sufficiently fine-grained yet persistent pain-state. If the mental state remains the same *at least as concerns* its mental properties, then the mental state can be said to be identical cross-time and *not* identical to any of its physical realisers. Thus, some token mental event could, and would, persist across changes distinct neural events.

One could once again assert that the event *in toto* is individuated partially by its physical characteristics, such that a shift in them generates a different mental-physical event. Thus the position is shielded from any sort of MR attack.

Certainly such a move is available, but I would suggest it might not be *desirable*, at least for those motivated by MR concerns. This is straightforwardly because whatsoever the metaphysics of events wind up permitting one to say about such cases of apparent multiple realisability, if we are motivated to do justice to the appearances here (and those persuaded by MR arguments against type-identity ought to be so) then we should say that there can be a mental state or event that is realised by distinct neural events. It simply makes more sense as a reasonable accounting of what we observe or imagine in such cases to deny mental-physical identity. We are confronted with a choice: either there are distinct events, one mental, and one or several physical, with some sort of intimate relation between them (or on a Davidsonian view (Davidson, 1970) the same event instantiating multiple properties, mental and physical), or one unified mental-neural event which cannot survive any physical change. Our stance on events in general should not be salient here, since we are beginning this argument agnostic as to whether the mental and physical events in question are identical in the first place, and the thought experiments suggest we should not assume so.

The former proposition seems intuitively to “answer” the challenge of MR in a way the latter dodges. This dodge may not be formally unjustified, though it ought to be unconvincing to anyone concerned with MR, unless a reasonable account can be given for just why we should suppose that the mental and physical cannot come apart. And, per Kim at the start of this section, whatever further account is offered should *not* extend to cases of type-identity.

This is not a cast-iron argument against the possibility of token-identity, of course; though one might think the available options do not look good in light of Kimean worries about the formulation of events for a token-account. There are significant and persistent worries for a token-identity account, and any view which relies on or incorporates token-identities as a load-bearing feature. Now, I consider realisation NRP. In Section 3.6 I offered a basic account of realisation, due to McLaughlin (2007), which holds that some property X realises property Y just in case X fills the causal role occupied by Y (ibid, p.151).

At first glance, absent some way to distinguish what “filling” is from what “occupying” is, it seems that we have two properties inhabiting the same causal role, and no reason to do so. Assuming something like overdetermination is a viable metaphysical principle and taking for granted given “causal role” that both are meant to be causal, we have overdetermination here. Unless, that is, these properties are identical. Certainly, some realisation-physicalists are, in fact, reductionists or identity-theorists of one stripe or another. However, we want a view that is non-reductive.

We can find elaborations in the literature, however. Stoljar presents a canonical form:

Realisation View: 'Physicalism is true at w if and only if every property instantiated at w is either identical to some physical property instantiated at w or is realized by some physical property instantiated at w' (Stoljar, 2010, p.123)

Understanding this formulation requires an introduction to second-order properties. On this view, every instantiated property that is not identical to some physical property is a *second-order property*. Take some glass to have the realised property of "fragility". This glass has the 'property of having a property' (p.122-3) that makes it fragile. Mental properties are likewise second-order on this view.

As Stoljar states; 'a property F realizes a property G if and only if F is the second-order property of having a property that satisfies certain conditions C, and G is the property that satisfies those conditions' (2010, p.123). G here is the mental property, where F is the base physical property. Yet condition C could, worryingly, involve 'being in a world that contains souls' (p.123) or some other condition wholly incompatible with physicalism. Left here, this formulation cannot define physicalism. Stoljar canvasses options available for the realisation physicalist. Taking the two I find most promising for physicalism quite generally, we have the stipulation that C is 'identical to something physical' (p.123), which is no good for our NRPist, or that C is itself realised, which leads us to infinite regress (p.123).

I have no space to consider the full suite of possible responses in favour of realisation; I offer this up only as a serious, perhaps non-fatal, concern for one prominent species of

realisation NRP, which predictably terminates with concerns about a pivot to identity- I elaborate on this emerging pattern later.

Something ought now to be said about the subset-realisation view, or set of views.

For Shoemaker-realization/the subset view;

It is a given for Shoemaker and Wilson that, while mental states do not reduce, their realization consists in the identity of the forward-looking causal powers of a mental property-instance being identical with the causal powers of a physical property-instance. This allows for the construction of an analogous MR argument for these properties, due to Pereboom (2002, p.133):

"Consider Anne's belief at some particular time that her parents live in Manhattan—a mental token, an instance of a mental property—and the token causal power [with] which it is identical. Suppose Anne is threatened with an illness that would damage a small part of her brain that has a crucial role in realizing this belief (but other parts of her brain have important roles as well). Before this part is damaged by the illness, a neurosurgeon could remove it and replace it with a sophisticated electronic microprocessor—let's call it a silicon prosthesis. Imagine that the illness never actually materializes, and Anne does not undergo the operation. Still, this token belief would have retained its token mental causal power had she undergone the operation, and had it thus at that time been realized by the token neural-and-silicon causal power instead of the token neural power that actually realizes it.

Consequently, the token mental causal power cannot be identified with a token neural causal power, specifically not with the token neural causal power that actually realizes it."

Stripping out the colourful example, the argument works similarly to that aimed against types, and token-events. The relevant tokened mental causal power could have been differently realised, and is thus non-identical to the realising physical causal power[82]. We once again see an attempt to develop a picture which involves some form of identity, and which is consequently vulnerable to MR argumentation. It should hopefully be clear, at this point, that *at least many* prominent NRP formulations share a common factor which leads them to similar sorts of dilemma.

In the introduction of this paper, I noted a difference between "thin" and "thick" non-identity, worth elaborating upon now. "Thin" non-identity is what I use to refer to formulations of NRP which attempt to cleave as close to identity as possible between mental and physical properties or events. This may be via identification of tokens, or via identification of forward-looking causal powers. Loosely, thin non-identity would pick out relations like the "non-identity" between the clay and the statue. Meanwhile "thick" non-identity would be cases of genuinely distinct entities like the non-identity between smoke and the engine producing it, or between Tony Blair and Theresa May[83].

[82] As noted, other sorts of realisation are available for the NRPist.

[83] This terminology is inherently clunky, since postulating "non identity" strictly seems like postulating the lack of a relationship. These are, rather, different ways in which certain entities might fail to hold the identity relation.

Thin non-identity requires some metaphysical relation between events or instantiated-properties, besides spatio-temporal or causal. Further, the relation must ensure the higher-level event is nothing over and above the lower-level event to which it is thinly non-identical. Two events are thickly non-identical, meanwhile, if they lack the aforementioned sort of relation and instead bear only spatial, temporal, or causal relations between them.

I intentionally omit any more stringent definition of thin non-identity in particular since, with this label, I intend less to pick out a particular technical or logical commitment, and moreso to capture a sort of approach or methodology. Thin non-identity is what I ascribe to the relations picked out in many NRP formulations which attempt to minimise ones commitment to anything beyond the physical by, as noted, cleaving as closely to identity as possible. Often, as we shall see, this manifests as building identity *simpliciter* into ones view somewhere as we have seen, or asserting relations which may well turn out to *be* identity, as I consider in the next section. Thick non-identity carries implications which may contravene standard-NOAA, intuitively. In later sections, I aim to show this is not necessarily the case, certainly not for "Nothing-Left".

Given the apparent failure of token identity, and of realisation-subset views, we are in need of a new formulation of NRP which avoids MR worries.

3.8 The Root of the Problem, or How I Learned to Stop Worrying and Love Irreducibility

Pereboom is instructive to cite in the previous section for a few reasons. Firstly, his analysis of the problem posed to NRP-theories here is in line with my own; he identifies that these

formulations of NRP involve identity in their characterisation of the mental-physical relation such that MR can be re-ran against these positions. However, his solution reflects an explicit commitment to standard-NOAA and thin non-identity which also leads his own solution to fail, albeit in quite a different way.

Pereboom proposes a form of constitution-based NRP. In particular, he favours the adoption of a distinctive non-identity relation between mental and physical properties to avoid MR worries, the "wholly-made-up-of" constitution relation. This is somewhat primitive, but, he argues, not obscurantist- we all know what is meant when [x] is described as "wholly-made-up-of" [y] (Pereboom, 2002, p.137). Further, we understand this relation to be asymmetric, and irreflexive (p.138). The putative benefit here is that constitution is ontologically innocent- it does not require the postulation of entities over and above the physical in any worrisome sense.

I do not take there to be an obvious construction of MR that could undermine this view. However, by my lights, there is another identity-based concern which emerges. Let us begin by considering what the relation is meant to "look like" intuitively.

First, we have the two descriptors already provided; asymmetric and irreflexive. These are standard parts of philosophic terminology, but I spell them out for sake of clarity. An asymmetric relation is any relation where x bears some property to y such that y cannot bear the same property to x. As a simple example of asymmetry, Rebecca may be taller than John, while John is not likewise taller than Rebecca. An irreflexive relation is simply a

relation which does not hold between a given entity and itself. Pereboom further notes that the relation he describes is ordered from less to more fundamental (2002, p.139).

Besides this, we can consider some illustrative non-mental example to which we might apply a "wholly-made-up-of" relation. Pereboom provides one of his own, noting that a diamond is wholly made up of the lattice of carbon atoms constituting it (2002, p.138). Certainly, it seems right to describe the diamond, in common-language, as wholly made up by the lattice. By my lights, this is coherent and readily understandable.

Pereboom takes this relation to be, at root, 'unanalyzable and thus primitive' (p.138), and basic in the sense that it cannot be reduced to or rendered in terms of 'more fundamental relations' (p.138). We should perhaps be sceptical of stipulated primitive or unanalysable relations, but this is not the core of my concern here. Rather, my worry is that the seemingly intuitive, basic, recognisable notion of "wholly made up of" to which Pereboom is appealing derives from our intuitive, recognisable, understanding of *identity;* spelled out further "x is wholly made up of y" is derived from our understanding of "x is identical with y". I contend that "wholly made up of" is not the most obvious reading or description we would provide, if prompted to describe the diamond-lattice relation. Rather, it is natural to think of the exact lattice structure as *being identical* with the diamond, such as it is.

In general, if I imagine anything which is wholly made up of its various constituents, then it, once again, seems natural and intuitive to consider an identity relation between those things, or rather, that thing. If I am wholly made up of, say, my biological body, my experiences, and anything else we might care to name which constitutes my person, then it

seems reasonable to say I am identical with that whole collection of phenomena. “I” come for free, after all in that set is present[84].

It will be worthwhile to consider identity itself, briefly. Identity is a relation shared between something and itself, it is ontologically innocent, reflexive, symmetric, and not ordered in any sense between more and less fundamental entities. The chair is identical to the chair, the Prime Minister is identical to the Prime Minister, and the black hole is identical to the black hole. Certainly we can assert that the black hole is wholly made up of the totality of its constituent parts, ditto for the chair and PM. Yet, how is this to be distinguished from the claim that the black hole, once more, is identical with itself, or with the totality of its constituent parts?

I am not, to be clear, asserting that no constitution relation is available, or that all necessarily reduce to identity. The car is not identical to its parts, but it does seem to be constituted by them. We may suppose that the car has differing persistence conditions to the set of its parts and therefore, that there is no identity between them by Leibniz’ law. Pereboom should not be understood to merely be asserting *some sort of constitution* is at play in the mental-physical case, but that we have a primitive and intuitively plausible relation at work, the same one at work in the diamond-lattice case. It may be clear enough that a constitution relation is rightly ascribed to the car-parts or diamond-lattice situations, and that identity cannot be so-ascribed on basis of persistence condition differences- it is less clear that this specific “wholly-made-up-of” relation applies. This is because a statement

[84] Certainly, if “wholly made up of” turns out to be a real relation distinct from identity, it would do the job of satisfying C1, in virtue of this.

"x is wholly made up of y" does not entail that x could persist beyond some change in y. If I state that the diamond is wholly made up of the lattice, or the car wholly made up of its parts, we may intuit that *that* diamond or *that* car would cease to exist beyond some change in lattice or parts, respectively. Pereboom presumably does not conceive the relation this way, but we are discussing an intuitive and primitive relation, and I see no decisive intuitive stance to take on persistence conditions in this regard; hence no clear delineation from identity.

More can be said, however. I note here a parallel with the debate concerning composition and identity. Lewis famously argued that composition *is* identity, though this involves some broadening of the identity concept (Lewis, 1991, p.84-87). Others go further in claiming composition is strictly identity (Baxter, 2001, for one example in the literature). For Lewis, at least, this takes the form of five "likeness claims" (due to Lewis) uniting the two; ontological innocence, automatic existence, unique composition, exhaustive description, and same-location (1991, p.85).

It seems to me these likeness claims are instructive for the relation under examination here. All these claims are true of, and central to, identity, and if they are true for the "wholly made up of" relation also, that should be cause for concern that this primitive relation is really deriving its intuitive force from identity. I argue the "wholly-made-up-of" relation likewise shares these qualities with identity. To begin with the most obvious, any [x] wholly made up of [y] will also inhabit the *same location* in space-time. The "exhaustive description" claim for identity runs that any [x] identical with [y], an exhaustive description of one will also be an exhaustive description of the other (p.85). This too, seems to run for

any [x] wholly made up of [y]. I cannot imagine a scenario where one could provide an exhaustive description of the whole set of constituents making up some entity which would not also constitute an exhaustive description of that entity. I covered ontological innocence earlier, and do not suspect that Pereboom would object to his notion being ontologically innocent, regardless. This leaves us with “automatic existence” and “unique composition”.

Unique composition is just that where [x] is identical to [y], anything identical with [y] will also be identical with [x] (Lewis, 1991, p.85). I think this goes for “wholly made up of” to. Take some [x] wholly made up of [y], anything identical with [x] will be wholly made up of [y], and anything identical with [y] will be wholly identical with [x]. If the diamond is wholly made up of its lattice, then anything wholly made up of that lattice will be identical with the diamond. Nothing in “wholly made up” supposes that the diamond could survive some change to its lattice. This differentiates this relation from supervenience. Loosely, one can substitute some constituent part of some supervenience base and preserve the supervening property. However, if we take seriously that [x] is wholly made up of [y], substituting some constituent part of [y] would make for a change in [x], or the destruction of [x]. Automatic existence is simply that wherever some [x] exists, anything identical with [x] also exists (p.85). It is trivial to see that wherever some token [x] is made up wholly of [y], since [x] exists, [y] will also exist. Therefore, Pereboom’s stipulated relation satisfies all the likeness conditions which led Lewis to identify composition with identity.

One may deny, by way of objection, some of these likeness claims as holding on the identity side. They do seem to form much of the canonical view, but this would be an option. I do not intend to address this, however. My focus is not on taking a firm stance upon how

identity should be understood, only noting that a significant part of the intuitive understanding we have of identity holds for the "wholly made up of" relation. Again, the relation is meant to be primitive, thus I would argue that we should be able to see, intuitively and pre-theoretically, that these hold for this relation, and thereby distinguish it. I am uncertain that this work can be done.

We should not simply accept a description of this purported basic primitive relation with addended qualities of "irreflexivity", "asymmetry", and the aforementioned fundamentality ordering. We can read these as stipulations which Pereboom makes of the relation to distinguish it from identity, but we have no reason to accept them, or the relation itself, as *distinct*.

Elsewhere, Pereboom construes his view differently, stating that a mental event token is wholly constituted by some physical event token when said physical event, and any further relational features, is 'sufficient' for the mental event's being tokened (Pereboom, 2002, p.500). However, as noted by Melnyk (2008, p.1292), this would seem to permit far more than what the "wholly made up of" notion does. In particular, it could very well allow emergence, contra C1;

> 'sufficiency might be sufficiency in accordance with a fundamental law of physical-to-mental emergence whereby, if an event of p's physical type occurs, then an event of m's mental type occurs' (p.1292)

Recalling Lewis and Baxter, I hope also to convey a broader issue for the set of views to which Pereboom's belongs, composition and constitution physicalism (Crane, 1995, p.3). For any formulation of NRP which appeals to composition, we have to consider the possibility that the notion they are using to capture the mental-physical relation may in fact just be identity. Without advancing those arguments now, I note that such positions seem unfortunately at risk of collapsing into some sort of identity physicalism, which was precisely what we set out to avoid in the first place[85]. In attempting to avoid MR issues, while cleaving to thin non-identity, we once again find collapse into identity, or MR's "revenge", across the views discussed in the last two sections.

Discussing every formulation of NRP would require a book-length treatment, far beyond the scope of this paper. Suffice to say, I take it, and have sought to demonstrate, that major formulations of NRP fall prey to problems stemming from multiple realisability, or their commitment to thin non-identity. In particular, NRP rejects mental-physical type-identity on grounds of MR. Certain NRP formulations are identity-involving elsewhere, identity between causal powers for instance, and mine and Pereboom's MR arguments can be ran against them. Token physicalism is identity-involving but supplants types for tokens; we have seen how this is also vulnerable to MR, if it does not simply devolve into type-physicalism in the first place. Finally, views such as Pereboom's do not explicitly involve or invoke identity but may nevertheless employ an identity relation *in disguise* as another, primitive relation.

[85] Parenthetically, I should not be taken as claiming that any constitution model is destined to fail as a solution here (though I'm sceptical of their capacity to deal with other problems presented to NRP, namely Exclusion); it is possible that such a theory could work out these kinks eventually. However, I think they are principally motivated by a constraint that is simply *too restrictive*. Minimally, to "succeed", such views will need to find a way apart from reduction.

3.8.1 Why do these views go wrong?

> "Still, given the truth of physicalism, there must be a sense in which the token causal power of mental state M would be "nothing over and above" the token causal power of microphysical state P—M's causal powers would nevertheless be "absorbed" or "swallowed up" by P's causal powers" (Pereboom, 2002, p.136)

As stated, it is not implausible that some variation upon a view I have thusfar described, or some other formulation I have not canvassed, which subscribes to thin non-identity in some form, may avoid the sorts of problem I have highlighted. I have not considered all composition or constitution formulations, for instance, and these may not face any sort of identity-related concern. Perhaps the search, as it were, should continue. However, we might ask why this should be thought a worthwhile undertaking in light of two considerations. First, the same issues are liable to be recurrent for any view committed to thin non-identity. Second, as I will argue, a better alternative is available, at relatively low theoretic cost.

To elaborate on my first point, views committed to thin non-identity are liable to either include some identity within their position, thus making themselves vulnerable to MR, as with supervenience and token identity when conjoined, and the subset view. Quite apart from this, they may fail to answer the challenge of the exclusion problem (Kim, 2005), which thin non-identity in conjunction with an acceptance of the causal efficacy of the mental seems to yield. I say more about how this might be avoided in an earlier paper, but my

position there, while accessible to the NRPist, rests on certain principles regarding causation they may not accept for other reasons.

It is clear enough that thin non-identity is motivated by a desire to satisfy standard-NOAA. Perhaps, however, we need not worry about this. NOAA is *somewhat* slippery, not rigorously defined or consistently understood, despite near-constant allusions. Our commitment to standard-NOAA should be tempered modestly, to Nothing-Left. This is because NOAA better sums up an attitude than a terribly precise principle; a desire to avoid ontological excess, or to "go beyond" what is permitted by the causal constraints of Closure principles. Yet, what Closure permits is, itself, debated[86]. Further, as I noted in the introduction, NOAA is seemingly part of the intuitive physicalist picture, and how physicalism is differentiated at first glance from any sort of dualism or panpsychism[87]. If what we really want, from NOAA, is to capture something intuitive, like the primacy of the physical in both instantiations and as a fundamental level of reality, and to avoid violations of closure, then what is left is to argue that we can get all of this, and more, out of Nothing-Left, which represents the fundamental core of standard-NOAA, and a sufficient part of the intuitive physicalist picture.

Suppose that we, *arguendo*, abandon standard-NOAA, or relax it to Nothing-Left. As a reminder, here is the principle we are left with:

[86] The most reasonable closure principles only tell us to not except there to be any physical effect without sufficient physical cause. Alone, they fail to even rule out downward causation.
[87] For dissent regarding whether panpsychism is differentiated from physicalism or is in fact the "real" physicalism, see Galen Strawson (2006).

"Take the physical away, and there will be nothing left".

I begin by noting, or recapitulating, what this principle permits, distinct from NOAA, before moving to consider what it still prohibits.

Most importantly, for present purposes, I take it that Nothing-Left permits thick non-identity between physical properties and mental properties, as well as other properties which may be multiply realisable such as, potentially, chemical or biological properties. Nothing-Left contains no prohibition against properties strictly "above" the physical. Rather, it permits them so long as they remain wholly dependent upon the fundamental physical properties for their existence. If there is a possible world where the mental properties can persist in the absence of the physical properties[88], then this would violate Nothing-Left. Of course, within the actual world, this principle would align with the notion that wherever some physical property ceases, any dependent mental property will also cease, in the absence of some other physical property-instantiation "replacing" the role of the physical property, as with temporal instances of MR.

It does also seem to permit emergence, when understood as higher-level or mental properties possessing 'causal powers which are not explicable in terms of the causal powers of their physical substrates' (Crane, 2001, p.210-11). That is, it permits them at least insofar as they could not exist or obtain without the physical properties from which they emerge. However, as discussed earlier, even NOAA was only part of the physicalist picture, one

[88] At least, if we take Nothing-Left as a necessary claim, though more on this shortly.

commitment of NRP. So too is some formulation of causal closure, which may itself prohibit emergence regarding downward causal powers, so perhaps we need not worry if Nothing-Left alone is insufficient. However, I do intend to address emergence in more detail in the next section.

As previously discussed in 3.6, specifically point (3), Baysan notes that one classic way of differentiating NRP from naturalistic dualism is modally;

> 'While non-reductive physicalism holds that the mental metaphysically strongly supervenes on the physical, naturalistic dualism holds that the mental nomologically (and not metaphysically) strongly supervenes on the physical' (Baysan, 2020)

I align with Baysan in taking it that naturalistic dualism cannot hold with metaphysical necessity, since this would eliminate the possibility of zombie or ghost worlds which motivate the position. Keeping this in mind, I will now briefly reiterate what this entails for naturalistic dualism, before elaborating on why Nothing-Left NRP can maintain metaphysically strong supervenience in contrast to naturalistic dualism and mere nomological supervenience.

There are metaphysically possible worlds, for the naturalistic dualist, where the mental does not supervene on the physical. For example, the naturalistic dualist will admit the metaphysical possibility of P-zombies, entities which are 'physically and functionally' (Chalmers, 1996, p.95) identical to us, while lacking any phenomenal consciousness. They may be ruled out within our world, for the naturalistic dualist, who admits that mental

properties actually depend upon physical properties, but they are nevertheless taken to be a real, non-nomological, possibility. That is, P-zombies and the like cannot exist in worlds which fall under the same natural laws.

I should note, if only parenthetically, dissent here. Lewis among others will take issue with this differentiation, not least because he regards physicalism as genuinely contingent, not a necessary, if a-posteriori, position. Lewis remarks '[physicalism] is meant to be a contingent thesis, a merit of our world that not all other worlds share' (1983, p. 362). If one shares this perspective I hope to demonstrate that the formulations I will soon propose can do enough to ensure the intuitive physicalist picture absent lining up with other sorts of NRP modally. My proposed formulations will be constructed to secure metaphysically strong supervenience, but at least one can accommodate contingent physicalism, as I will demonstrate.

That aside however, the NRPist might feel more confident about accepting Nothing-Left if it can be shown as sufficient to reject the above modal view on the nature of mental properties held by the naturalistic dualist. I think it does just that, at least so long as we take it to hold with the same necessity that standard-NOAA does. For any world, where there are no physical properties, there can be no mental properties "free-floating", or "P-ghosts", on a standard reading of Nothing-Left. We rule out, easily, any sort of mental "ectoplasm" or phantasma, for a start.

While P-ghosts are easily ruled out by Nothing-Left alone, P-zombies do not seem to be so. This is remedied when we remember that NOAA, and Nothing-Left are, again, only part of

the intuitive physicalist picture. We might say that the formulation of Nothing-Left I initially offer needs to be understood as a companion to Dependence, yielding something like this; "every instantiated property is dependent upon (or identical to) some physical property, take the physical away, and there will be nothing left". Here, then, it becomes clear that Nothing-Left, conjoined with Dependence, clearly only permits views with the modal view of mental properties shared by NRP, not naturalistic dualism.

I do not think it *ad-hoc* or unreasonable to pair Nothing-Left with Dependence here to fend off this dualist concern, given the close intuitive and conceptual relationship between the Dependence thesis and the Nothing-Left principle- or rather NOAA, of which Nothing-Left is a constituent, as I outlined early in this work.

One might reasonably ask what work Nothing-Left is doing here, if Dependence can "save us" from any sort of slide toward naturalistic dualism. I think the main factor here, for the intuitive physicalist or NRP picture, is that Nothing-Left secures the primacy of the physical in our ontology. In the actual world, not possible worlds, Nothing-Left makes clear it is the base physical properties from which everything else comes, and in the absence of which, one is left with nothing.

To return to non-identity, Nothing-Left allows us, I think, to consider relations between the physical and higher-level properties which are not "thin" non-identity. If some physical property is causally necessary for the synchronic emergence of some non-physical property, we can still say that removing the physical property means one will be left without the non-physical property as well. Yet causation is clearly going to relate entities of any kind in a

non-thinly-identical way. I am thickly non-identical with some property I have caused, even when that thing may *depend upon me* for its existence. Quite simply, we can allow *genuinely irreducible,* mental properties, grounded by physical properties, without violating the naturalistic spirit of Physicalism. Without the microphysical properties, there are no other properties.

However, we cannot accept just any thick non-identity relation between mental and physical properties. What is needed is something which satisfies Nothing-Left, *along with* the other theses and commitments of NRP. Fortunately, there are at least two candidate relations available to us here.

3.8.2 Grounding as Metaphysical Causation, or Synchronic Causation

There is a vast literature on grounding, understanding grounding in a host of ways, some quite notably distinct from others (Rosen, 2009, p.109-112). I do not intend to wade into that debate, nor to declare my preferred notion the "only" or "best" way to understanding claims like "x exists in virtue of y" or "x is grounded in y". I note here only that Alistair Wilson's formulation of grounding, as metaphysical causation, seems a coherent and reasonable relation, and one which is likely to be of use to us here. Henceforth however, to avoid confusion, I refer to Wilson's grounding as W-grounding[89]. Perhaps other grounding

[89] Importantly, I focus on Wilson's version of grounding not because it allows me to provide a grounding account of mental-physical relations, but because it permits synchronic causation. Wilson's grounding is causal, and I take it as a fairly strong shared intuition that causation is distinct from any sense of identity. If x causes y, x is not identical with y. Thick non-identity "comes for free". Our typical cases of causation have the shape "I kicked the football, which caused it to roll across the ground". W-grounding also gives us a way of explaining the essentially derivative nature of the mental while maintaining its irreducibility.

relations ought to be part of our best metaphysics, or perhaps not, but this need not concern us here[90].

'Grounding is a way of causing' (Wilson, 2018, p.723)

To begin, Wilson considers our usual notion of causation "out in the world", namely 'nomological causation' (p.723). The throwing of the stone which causes the glass to break, 'mediated by the laws of nature' (ibid) is nomological causation. On the other hand, 'the existence of Socrates is a sufficient cause of the existence of Singleton Socrates is not mediated by any law of nature' (p.724). Singleton Socrates exists in virtue of, or is grounded in, the existence of Socrates. Wilson notes two benefits of considering this grounding to be metaphysical causation; theoretical unity, and parsimony. The view is parsimonious, since we need invoke no new 'primitive notions' (ibid), to explain relations in our ontology. We already "have" causation, howsoever one understands theory of causation, and can do theoretical work with it. I have already objected to offering up wholly new relations for NRP within this work. Second, we typically understand grounding claims as explanatory, and this allows us to account for that. Causal explanation identifies nomological causes for events in nomological causation, and grounding is deployed when identifying metaphysical causes, in metaphysical causation (ibid). One final consideration, independent these benefits, for why we might regard W-grounding as a species of causation is the strong analogy between full and partial grounds and sufficient and contributory causes (p.747). A sufficient cause or a

[90] Wilson does take his view of grounding to be the way that any and all "grounding" relations ought to be understood (Wilson, 2018, p.723), but nothing rides on this for our purposes. In addition, it should be noted that certain construals of grounding would seem to fall under the banner of thin non-identity, and hence should be ruled out as appropriate for the mental-physical relation by those who have found my argument thusfar compelling. See Rosen (2009) for further discussion on this point.

full grounds are sufficient for the caused or grounded fact, respectively, to come about. Likewise, partial grounds and contributory causes are only partially sufficient for the relevant fact.

These benefits, at least the first, do hinge upon taking “grounding as metaphysical causation” to be the “only” grounding, and though I do not wish to take a stance there, one might find such arguments reasonably persuasive.

Metaphysical causation does not rely upon the positing of any new primitive relations (p.724-5), and it clearly allows that the grounded may be thickly non-identical from their grounds. Causation, and hence metaphysical causation, is asymmetric and irreflexive (p.728). One note of caution here would be that Wilson does identify the relata of causation, and hence metaphysical causation, as facts (p.729), not property-instantiations or properties, as I have been doing through this paper. While discerning the relata here is important for any position, I do not think that much rides on it for present purposes, for the following reason. If causation, and metaphysical causation, relates facts rather than properties, this does not do away with any of the MR concerns highlighted for positions earlier, which are not premised upon anything to do with the metaphysical categories of the physical or mental “states” in question. If, on the other hand, it turns out that causation relates properties or property-instantiations, or that views taking facts to be states of affairs made up of property-instantiations (Armstrong, 1986, p.85), then we can talk of the same for metaphysical causation.

Another element of W-grounding seems to fit perfectly into the intuitive physicalist picture, regarding the primacy of the physical, Dependence, and physical fundamentality. As Wilson notes, grounding in general is taken to say something about fundamentality, where the grounds are more fundamental than what is grounded (Wilson, 2018, p.731). One serious issue for his view is that *nomological* causation does not relate its relata in terms of fundamentality, and given this we might worry that metaphysical causation similarly lacks this feature.

First, nomological and metaphysical causation are distinguished via mediation by laws of nature, as previously noted, where this goes for nomological causation, not metaphysical. Thus;

> **'**dependencies between facts about different levels of reality are typically not mediated by the laws. This suggestion explains why nomological causal connections typically structure the world with respect to time, while metaphysical causal connections typically structure the world with respect to fundamentality' (p.731)

Where nomological causation structures reality through temporal relations between relata as mediated by laws of nature, metaphysical causation structures reality through differences in fundamentality between its relata. For W-grounding NRP, the fundamentality of the physical is a clear part of its relation. We can even say more, where we understand reality to be ordered from the microphysical, to the chemical, then biological, then mental. W-grounding embeds a fundamentality-ordering which allows us not just to situate the mental as less fundamental than the physical, but to theorise about differences in

fundamentality between levels, some of which may also feature multiply realisable properties.

W-grounding offers us a thickly non-identical, causal relation between physical and mental properties (I return to property talk for ease, though as noted I do not take much to ride on this). It ensures Irreducibility and Dependence, minimally allows for us to satisfy C2, since nothing in W-grounding implies that grounded properties may exert closure-violating causal influence. Further, it captures Nothing-Left's intuitive appeal since once you have the physical properties, you have all the other properties *in virtue of* having the physical properties. To recall a point made in my introduction regarding NOAA;

> 'Mental facts "get to exist" derivatively, in virtue of fundamental physical facts, mental properties in virtue of physical properties, and so forth: 'non-reductive physicalism must **honor the intuition that fundamental physics is the basic science**, [] (Melnyk, 1995, p.371, emphasis mine).'

This all remains true with W-grounding. However, does it keep us on the "right side" of the modal gap between NRP and naturalistic dualism, as outlined earlier?

I think it does. While Wilson remains agnostic on the question of whether full grounds necessitate what they ground, he allows, alongside many other grounding-theorists, that they *may* do so (Wilson, 2018, p.731). Further, if our laws of nature are contingent, meaning that nomological causation does not ensure the necessitation of effects by causes, this is no concern for metaphysical causation regarding necessitation since metaphysical causation

does not involve laws of nature, contingent or no (Wilson, 2018, p.731). We can then reasonably assert that W-grounding of mental properties by physical properties is metaphysically necessary, and consequently holds in all possible worlds, *contra* the naturalistic dualist.

It is sometimes natural to think of grounding as a reductive relation. In particular, grounding can be understood as a building or 'constitutive' (Goff, 2017, p.43) relation such that the grounds simply constitute the grounded. For an example, the party may be grounded in the individuals standing around chatting. The party is nothing over and above these people, in a clear sense. However, we need not regard all grounding relations, especially not the one we are considering here, as constitutive. Goff mentions Moore's ontological characterisation of goodness. This was taken by Moore to be fundamental, though nevertheless supervenient upon non-moral facts as I mentioned earlier. Goff then suggests we could regard this as a 'non-constitutive grounding relation, in which the facts about goodness are grounded in but ontologically additional to the non-normative facts' (p.43). Therefore, I must note that my W-grounding is non-constitutive grounding, and I already accept that it represents some "ontological addition". Finally, I do not think this is out of line with Wilson's own view, considering the non-constitutive nature of some of his own examples of grounding:

'Part/whole: The existence of my head grounds my existence.
Act-consequentialism: Act A's having the best consequences grounds A's being right.
Euthyphro: God's desiring that P grounds its being good that P.
Noether: The symmetry of the laws of nature under time-translation grounds the fact that energy is a conserved quantity.' (Wilson, 2018, p.732)

One may object to my offering up W-grounding as our solution here. This is not a well-established relation in the literature, even if it may become so. We might see objections to the relation itself, and my view ought not to be hostage to fortune regarding metaphysical relations. I take W-grounding to be the optimal relation up for grabs here, but there is another option that seems to fit theses and commitments I have been concerned with.

I have in mind "mere" causation itself. However, this says too little. We need to refine the sort of causal relation in question. I begin by noting the need for a realist understanding of causation, where causation is not merely an inference we are making about the world, but causal relations are real metaphysical relations. If one disagrees here, one may simply "get off the bus" with regards to my argument, though I take it that anyone with that view would have already gotten off the bus, whether for C2, or for appeal to causation within various arguments already made. Productive, or generative, theories of causation hold that causation involves some sort of transmission, exchange, or production (Raatikainen, 2018, p.27, 31). I do not intend to argue for productive causation here, though for any NRPist who adopts such a view, the following may be of interest.

Productive causation does not require, of necessity, temporal relations. That is, it need not be the case that a cause exists at t1, however temporally fine-grained, and the effect exists at t2[91]. Certainly much causation does run "forward", temporally, and there may be metaphysically possible time-travel scenarios where some cause triggers a prior event. Yet,

[91] Though, see Wilson (2020, p.46)

none of this rules out the possibility of synchronic productive causation. I would not want to suggest I have strong independent reason to rule the possibility in, however I see no good reason against it and it has utility within this metaphysical framework.

Suppose we allow for mental properties being productively, synchronically caused by physical properties. This certainly looks close enough to naturalistic dualism that we have cause for worry. Taking Nothing-Left seriously, we need only stipulate one further condition upon this causal relation to bring us back into line with the intuitive physicalist picture, as I see it[92]. I call this "Dependent Causation"

Dependent Causation: For any property [x] synchronically, productively caused by some other property [y] (within the relevant class), [x] can only exist where:

(a) some other property [y] (within the relevant class) exists, and

(b) will necessarily be caused where [y] does exist.

I note "relevant class" so as not to invoke some sort of causal dependency between specific tokens. [x] here could be caused to exist by many different token [y]'s, but it could not exist without *some* [y] in the relevant class[93], and when [y] ceases to be, so too will [x].

[92] Close enough, at least. My next section gives further reasons to think the benefits of adopting such a formulation might outweigh concerns here.

[93] Recalling MR, the relevant class would be the total set of physical properties, any of which are sufficient for the mental property in question.

This stipulated condition may appear *ad-hoc*. In response to this, I would appeal to our basic understanding of strong emergence, whether or not one regards strongly emergent properties as possible. Strongly emergent properties are typically understood as synchronic, caused, and dependent upon their emergence-base (Melnyk, 2006, Rueger, 2000). For a causal relation which is similarly synchronic, I think it would be reasonable to import our understanding that such properties are dependent upon their bases. As for the necessity within Dependent Causation, I take this to fall out of our prior understanding of bases and realisers as necessarily co-existent, which comes out of Dependence.

I want to briefly turn attention here to the notion of strong supervenience, as provided here by Kim:

'Mental properties strongly supervene on physical/ biological properties. That is, if any system s instantiates a mental property M at t, there necessarily exists a physical property P such that s instantiates P at t, and necessarily anything instantiating P at any time instantiates M at that time' (Kim, 2005, p.33)

Strong supervenience between mental and physical properties is often supposed insufficient for physicalism, and though other reasons for this exist in the literature, I provide one relevant considerations here. Supervenience only picks out the co-occurrence but does not say anything about the relation itself between these properties. With dependent causation incorporated into this view, we establish a supervenient structure between mental and physical properties, satisfying Nothing-Left. The Causal View, however, additionally provides

a throughgoing relation *explaining* this co-occurrence, by positing a causal relation between the properties.

The picture that emerges is as follows. Physical properties synchronically, productively cause mental properties, and this relation is constrained by "Dependent Causation". I call this the Causal View.

With the Causal View, we have another relation which is readily understood, preserves Irreducibility and Dependence, *can* preserve C2, and which satisfies Nothing-Left. By ensuring, via Dependent Causation, that there can be no world where the causing property can fail to exist without the caused property, nor where the caused property can exist without *some* causing physical property, we guarantee that Causal-NRP ensures the supervenience of mental properties upon physical properties within all metaphysically possible worlds; we can still say that necessarily wherever some mental property is caused by some physical property, that mental property depends upon that physical property. The mental, here, only exists as a *causal* derivative of the causal-base physical properties, so the primacy of the physical is assured. Here too, for reasons of physical primacy, we may also stipulate that these synchronically caused mental properties may not, given their nature as causal-derivatives, possess novel downward causal powers of their own[94].

[94] This feature is a reasonable stipulation, I think, though it is non-essential. The aim here is to have a relation which is clearly distinguished from emergence but as Pereboom notes of his own formulation, 'ruling out emergence would require a separate condition on physicalism [and] no one, to my knowledge, knows how to rule out emergence by way of a more fundamental condition on physicalism' (Pereboom, 2011, p.137). Though I proceed on the assumption that this Is a reasonable aspect of the relation I outline, my failure here would only "bring me down" to the level of many other NRP-formulations in this regard.

Before moving on, something should be said here about contingent physicalism. The view, as previously cited from Lewis, wants to preserve the idea that the grounding of mental, and other higher-level properties, in physical properties is a special feature of our world that others may not share. The W-grounding relations I invoke involve necessitation, though *perhaps* this could be extricated. More promising, however, is the Causal View. Although dependent Causation means that my Causal View is necessitating, the contingent physicalist who is otherwise sympathetic could modify this constraint to avoid necessitation. Rather than "[x] will necessarily be caused where [y] exists", they might state "[x] will always be caused by [y] in the actual world". They may also wish to say that both formulations are acceptable *as is*, taking there to be some metaphysical principle which modifies the base physical properties within some worlds or which functions as a defeater on the causal or generative powers of these properties, such that these relations will not obtain without contradiction with the necessity of the relations generally. I do not know what such a principle could be, but understanding might be true in our world and false elsewhere is a debate I leave to contingent-physicalists. It does not seem to me, at any rate, that my proposed formulations would be worse off than most others in this regard.

Ontological innocence is another concern of the physicalist, recalling discussion of NOAA. While the causal view abandons some ontological innocence in terms of admitting mental properties as distinct, caused properties, it can be argued that the view is far more innocent with regards to the relations being posited. We no longer need to posit some sort of thinly non-identical relation, whether "realisation", or the wholly novel "wholly made up of" relation. We need only identity, and causation, to which many here are already committed.

Other relations may exist or be worthwhile ontological inclusions, but they are not *requirements* of this formulation.

A response to this might run as follows. Common sense or intuition already seems to recognise some relation between diamonds and lattices, cars and their parts, so forth, the relata of which hold differing persistence conditions. There may be no name for this common-sense relation *per se*, but it seems to be synchronous, non-causal, NOAA-satisfying, and *not identity*. This is just "how we think" about these relations. We can grant that common sense seems to desire some relation which vaguely tracks with these theoretic qualities, but it does not seem to me that common sense can get us this far. Once common sense has pointed us to something, we are still in need of philosophical work to extricate more precisely what is going on, and it is far from obvious that one could precicify the relevant common-sense notion in such a way as to make clear that the relation is non-causal. If we asked a 16th century stonemason whether his sculpture was *caused* to exist by the shape now taken by well-carved stone, is it clear the answer would reject causation? Further, is it clear this would go for the mental-physical relation which concerns us?

With the common-sense concern aside, we may still wonder whether this formulation does enough to preserve the intuitive physicalist picture overall, for the physicalist. Trogdon has this to say about NOAA regarding the mental, and its attendant ontological framing:

'so mental properties are nothing over and above physical properties in the actual world. But how can one property be distinct from, yet nothing over and above, another?'(Trogdon, 2009, p.148)

‘First, a property Q is actually nothing over and above property R just in case the actual world is such that if Q is instantiated, it is instantiated in virtue of R’ (p.149)

‘Central to sparse ontology is a hierarchical view of reality; the basic entities form the sparse structure of being, while the derivative entities form the abundant superstructure’ (p.147)

Trogdon is not atypical in understanding NOAA as involving some sort of “in-virtue-of” relation, and in taking NRP- sparse ontology physicalism with a commitment to Irreducibility- to involve base-entities and derivative entities forming a sort of superstructure above the base, which only exists in virtue of those base-entities. Both formulations in this section capture the derivative, dependent nature of mental properties through accepting Nothing-Left. In addition, they can accommodate the intuitive picture of a superstructure which can only exist in virtue of the basic physical properties. W-grounding in particular is rooted in capturing the intentional use of “in virtue of” in philosophical discourse.

This may not be enough for some, however. I note that these formulations fit the theses and outlined commitments, and have hopefully thereby demonstrated the ways in which constituent parts of the intuitive picture can be satisfied here. However, a full-blooded causal view may just be a bridge too far for many NRPists who would otherwise want to accept this sort of formulation. Perhaps many would be satisfied with W-grounding, regardless.

I offer one last, related, worry for my proposed formulations:

> 'Should we distinguish the materialist for whom the phenomenal facts are grounded in the neural facts together with contingent psychophysical laws from the dualist for whom the phenomenal facts are merely caused by the neural facts according to psychophysical laws? We should. The difference will be clearest if the dualist allows that the relevant causal laws may be indeterministic, for in that case the underlying facts will not necessitate the phenomenal facts. [It] strikes me as much more natural to keep causal relations on one side—as external relations among wholly distinct states of affairs—and grounding relations on the other' (Rosen, 2009, p.118).

Although Rosen is operating on a different view of grounding to that I have put forward with W-grounding, we might still be concerned that appeals to causation are wont to force W-grounding into the dualist "camp". I take the potential issue here to be that where some synchronic causation, common to both views, is *the* relation between physical and mental properties, then the physical properties will fail to necessitate the mental properties given that the causal or nomological laws connecting them will be contingent. I think this objection misses the mark. While the dualist may propose a synchronic causal relation which is contingent and may fail to hold across all metaphysically possible worlds, we are not obliged to hold that this is true for our synchronic causal relations. W-grounding is already necessitating, and I have already stipulated that the Causal View can be as well. Further, Wilson is explicit that his metaphysical causation need not involve laws of nature in the way we might think diachronic causation *does*. The division between NRP and dualism here is not between causation and some form of dependence or grounds, but between the

view one holds regarding the contingency of psychophysical laws. I think both of my formulations can join other formulations of NRP in rejecting the possibility that laws could be such as to allow for P-zombies, inverted spectra, and so forth.

My next section will show that, to the extent that one is concerned about some small departure from the intuitive picture here, the theoretical benefits can far outweigh this cost.

3.9 The Benefits of Thick Non-Identity, and Where We Go From Here

W-grounding and the Causal View are both significant departures from other NRP formulations. This section demonstrates that both, in their own ways, can earn their keep regarding several matters important to the NRPist. Aside from the ontological innocence of the relations involved in the formulation, as already discussed, I consider three such benefits, "Exclusion and Mental Causation", "Undercutting the Phenomenal/Psychological Divide", and "Accommodating Emergence".

3.9.1 The Exclusion Argument and Mental Causation

I have offered my answer to the Exclusion Argument at length in my second paper within this thesis. The worry, though, is that in denying the identity of mental and physical properties, the mental is robbed of causal efficacy. Kim begins his formulation of the problem with the case of mental-to-mental causation. To do so, he first appeals to "Edwards' Dictum", saying 'there is a tension between vertical determination and horizontal causation' (Kim, 2003, p.153). Some "later" mental property is "vertically determined" by

some base physical property, howsoever that is understood by the NRP-formulation in question. Meanwhile, per mental causation, it is taken that some prior mental property should *cause* that later mental property.

To underline the apparent tension, Kim appeals to an example proffered by Edwards. We imagine a world where God is the sustaining, immanent cause of all properties, at every time-slice. For a given time-slice, all the macro-properties we observe are vertically determined by God's placement of the micro-properties, with no work left for causal relations between properties across time (Kim, 2005, p.37). Returning to mental properties, all the work of bringing about some later mental property is exhausted by its microphysical base, leaving prior mental properties with no causal work to do. The Dictum motivates a more formal overdetermination principle, such as:

> 'No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination' (p.42)

Where the base physical property is sufficient for the mental property, no prior mental property is needed, and postulating both seems to invite apparently troublesome overdetermination.

Briefly, my response was reliant on certain causal principles, centrally difference-making as a principle of causation, and an understanding of causation at levels such that one could defend the possibility of mental properties causally relating to each other without their causal power "draining away" to their physical bases. Besides this, I also undercut the

motivating intuition for accepting this causal overdetermination principle, largely by demonstrating a substantial disanalogy between the case of mental causation and the scenario which motivates Edwards' Dictum. I do take this answer to be available to even NRPists who endorse views premised on thin non-identity. However, understanding mental properties as thickly non-identical to their physical bases provides a picture which more clearly "protects" intra-mental causation from any sort of drainage concern.

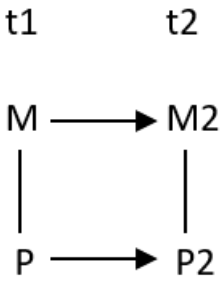
In that paper, I argued that NRP could preserve mental causation through a combination of endorsing difference-making as a principle of causation, and through incorporating behaviours as a macrophysical properties which nevertheless feature at the psychological or mental "level" of reality and which are subsequently commensurable with mental causes rather than prior microphysical causes.

To elaborate, the generalisable view for NRP would be that the proximate cause of some given mental property will be some former mental property, generated by some base physical property. This is because the prior mental cause is commensurable, or proportionally difference-making (List and Menzies, 2007, p.9-12, building on work from Yablo (1992)), to the mental effect in a way the physical base is not. What makes the causal difference for some mental property at time t2 is some prior mental property, not some microphysical property which is too *specific* (p.7). We know, given MR, that the microphysical base could have been otherwise and the mental property at t2 could still have come about. Meanwhile, in appropriately similar close possible worlds (ibid), the latter mental property would not have come about without the former. On either a W-grounding

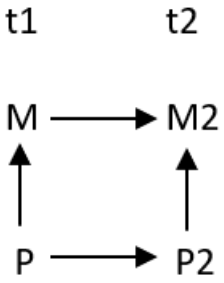
or Causal View, the mental-to-mental relations may still hold, but all relations in the network become causal.

We have two different networks that can be developed (non-arrow lines mark non-causal dependence, arrow-lines mark out causation):

Thin NRP



Thick-NRP



While I have not mentioned "preserving mental causation" as a commitment of physicalism, since it is not a germane commitment for much of this paper, I do take it that NRPists and

physicalists broadly wish to preserve mental causation. Thus, it is a good-making feature of any NRP-formulation that it can protect the possibility of mental causation.

Some unique concerns do emerge for my formulations specifically, relating to both the causal overdetermination principle and the difference-making principle. Principally, though I preserved causation for the mental by undermining the notion that there was competition between causes and determining-bases, this might not be thought to extend to undermining competition between mental and physical *causes*. Further, if proportional difference-making is a principle which allows us to state that some prior mental cause is a better proximate cause, then what do we say about the purported physical cause on either of my proposed formulations? It does not seem commensurable to the mental property in the same way some prior mental property *does*.

To the first worry, I have already undermined causal overdetermination principles across my other papers within this thesis and refer to them for in-depth discussion on the topic. One important insight, as I see it, is that not all overdetermination is created equal (Funkhouser, 2002, p.337). “Independent” (ibid) overdetermination, two disparate causes converging coincidentally upon the same effect where either would have sufficed, certainly seems worryingly coincidental. Where overdetermination is non-coincidental due to resulting from some prior shared cause, we have little reason for concern. This seems true of M at t2 in the thick-NRP network. Beyond this, though, we might think that the intuitive idea underlying the “weirdness” of ubiquitous causal overdetermination comes out of examples of cross-time overdetermination. A man being shot and struck by lightening at the same time has his death overdetermined by diachronic causes. We do not seem to have similar intuition

pumps for cases where the "overdetermination" occurs with one synchronic and one diachronic cause. To the extent one remains worried about causal overdetermination, it seems one ought to provide an argument for extending this to synchronic causes, which are already noted to be a different sort of causal relation in various ways; temporally, with reference to natural laws and necessity, and in terms of conditions by which effects come about from causes. We need not *assume* competition here, certainly if we are already compelled by arguments to reject competition between other sorts of dependence or determination relation and diachronic causal relations.

What of proportional difference-making? The principle holds that causes ought be proportionate to the effects they generate. If a pigeon is taught to peck at red objects, and pecks at a crimson object, it is the property of redness which is causal in this instance; crimson is too specific to be the difference-making cause, even though it is causally sufficient for the effect (List and Menzies, 2007, p.9)[95]. By employing this principle to establish a former mental property as the proximate cause of some later mental property, we seem to exclude determining physical *causes*, given they will seem non-proportional to the mental property[96]. Even absent my hope that my former model may incorporate thick-NRP, the difference-making principle is more broadly compelling as a constraint on causation and I take it many philosophers would have good reason to accept it, or something like it. So, we need to make sense of difference-making in relation to thick-NRP.

[95] This is explained in greater depth through counterfactual examples, but I avoid this elaboration for sake of brevity. Nothing in what follows rides on those specifics.
[96] Notably, this principle does not rule out joint causation or overdetermination. I take it that we can help ourselves to as many causes as we wish on difference-making alone, just so long as the causes are proportionate to their effects.

Fortunately, two non-exclusive responses can be provided. Firstly, synchronic causation, both in W-grounding and the causal view, is distinguished from nomological or diachronic causation in several ways. There seems no immediate reason in principle why difference-making should be applied to synchronic causation. One reason to motivate adoption of such a principle would be the construction of cases analogous to the pigeon-pecking case, for synchronic causation, with the same intuitive force. While we *can* apply that compelling reasoning from the pigeon-pecking case to *diachronic* mental causation, it does not seem that we can readily construct such cases to demonstrate a synchronic proportionality principle. Even if creative examples become available, it seems acceptable to deny the entailment of a substantive causal principle in the synchronic case while permitting it in the diachronic case. This is because it is simply *unnecessary* for synchronic causation. One noted difference between synchronic and diachronic causation is the former *necessitates*[97] while the latter is contingent (Wilson, 2018, p.730). The pigeon, as it were, may fail to peck. In the diachronic case, we are thus in need of principles to appropriately pair causes with their effects. In the synchronic case, on either of my formulations, a necessary relation between cause and effect is part of the model.

Beyond this, we can accept a "soft" difference-making principle which does not demand that causes be proportionate to their effects, but only that proportionality identifies one exemplary causal candidate, perhaps even a cause we *must* accept. This would not rule out other causes, even where not the "difference-making" cause[98]. The pivotal point is this does

[97] As noted, the contingent-physicalist *may* deny this necessity in adopting one of my thick-NRP formulations, but I take this move to be amenable to others.

[98] On the view I propose here, we could think of difference-making as sufficient for some property to be a cause, but not *necessary*.

not rule out causal efficacy for the microphysical property. This refinement may be independently welcome. We have other reasons to postulate causation and causes bar difference-making, such as inference from the typical causal powers of some property and correlation with a relevant effect, or the possibility of worlds where some non-proportionate property obtains absent *any* "difference-making" cause. We might imagine some world where God creates only one property; a microphysical property which is one of the "realisers" of a suitably specified pain state necessitating, causally or not, that pain state. Take my thick-NRP model to hold here, and *arguendo* stipulate a "hard", or traditional, difference-making principle. Thick-NRP would suggest that the pain is caused by the microphysical property. The otherwise-physicalist endorser of hard difference-making may not wish to deny the pain, on threat of dualism, yet could not accept a causal relation between it and the microphysical property. This leads us back to postulation of thin-NRP relations and their issues. We can just as easily say that such possible worlds demonstrate the synchronic causal relation, while nevertheless admitting prior mental causes in the actual world as principal causes of later mental properties.

Besides that, examples provided by List and Menzies (2007) provide good reason to think that difference-making marks out causes, but not that difference-making causes rule out all others. We should be metaphysically "modest" in our formation of causal principles, and this refinement to "soft" difference-making ensures we do not over-extend.

Kim has recourse; he may assert in the "thin" case that some physical property at t1 is not causally but at least *generatively* sufficient for the bringing about of some mental property at t2 (Kim, 2005, p.19) on thin-NRP. This is blocked on any thick non-identity NRP. The

mental property at t1 is part of the causal chain leading to the mental property at t2, and as such is "on all fours" in causal terms with the physical property causing it at t1. We do not generally take it that causes earlier in some causal chain undermine or drain away the causal powers of succeeding causes and should not therefore do so here. We must accept that both the mental and microphysical properties are causes of the mental property at t2. This is complementary with the rest of my argument in that paper, which establishes the possibility of causation between irreducible mental properties.

Supposing that NRP quite generally can answer the challenge from causal exclusion, at least through same-level causation as I argue, then we might think it entails something substantive about the status of mental properties. Consider the following principle:

Alexanders Dictum: "To exist is to have causal powers" (Cargile, 2003, p.143)

Alexanders Dictum does not receive universal assent from philosophers, but it has some measure of support, at least when it comes to concrete entities, properties and so forth. Let us take it to be true, for a moment. Having established that mental properties have causal powers[99], the Dictum would compel us, naturally, to admit that these properties exist. But this is trivial, of course. No-one would argue that some object, event, or property which had causal effects did not exist. Additionally, we must accept Irreducibility if we are to be NRPists. Therefore, it cannot be said that all that exists are properties identical to the base physical properties. Thin non-identity *at best* allows us to say that mental properties "exist"

[99] Not downward, only at their level.

in some way which is nothing over and above the physical. Yet this *wholly* derivative existence does not seem to do justice to the Dictum. Thick non-identity formulations meanwhile identify these causal mental properties as substantive, throughgoing existents, just like their underlying physical properties. Per my prior papers model of mental-to-mental causation, we might be especially compelled to view these causes, with distinct higher-level causal powers, as distinct existents in their own right. Though, as demonstrated in section 7, this alone should not give us reason to worry about our NRP bona-fides.

3.9.2 Undercutting the Phenomenal/Psychological Divide

Kim was a leading proponent of reductive physicalism for many years, and he remained concerned that qualia might prevent us from getting a total physicalism, only '[physicalism] near enough' (Kim, 2005, p.6). Kim wants to maintain the causal power of mental properties, and was thus understandably reluctant in accepting epiphonemal qualia (p.197) as irreducible. In particular, Kim took it that inverted-spectra cases gave reason for supposing at least some qualia were epiphenomenal (p.172-3), given that they seemed to make no causal difference and hence resisted functionalisation or reduction (p.172).

Further, concerns about the possibility of P-zombies leads Chalmers, among others, to propose a divide between the phenomenal and psychological, where psychological properties may be functionally or physically realisable, and phenomenal properties, or qualia, are not (Chalmers, 1996). While many physicalists take qualia to be something which can be, or may be, successfully reduced, or that challenges to physicalism from qualia-based

arguments might be answered, my proposal allows us to wholly eliminate this “spooky epiphenomena” concern.

On W-grounding, we can say that qualia are metaphysically caused just the same as any other mental property or event, ditto on the Causal View. Whatever arguments for the irreducibility of qualia may be brought, irreducibility is already accepted on these views, in as strong a sense as the dualist may wish to argue, except coupled with a substantive dependence upon their base causal physical properties. We have a framework which can then accommodate irreducible qualia, without compromising the primacy of the physical!

Per 8.1, they can also have causal impact, and per W-grounding, feature in causal explanations (Wilson, 2018, p.747). This is a further boon to NRP in terms of naturalist bona-fides, since we need not accept any sort of non-causal ectoplasm, nor cleave phenomenal properties apart from psychological properties within our scientific psychological theories, given that phenomenal properties are deployed in our psychology as practiced. Outside scientific practice, this seems to also help us comport with common-sense understanding of our own minds. Beliefs, intentions, and desires, seem informed by phenomenal states, and segmenting these states seems counter-intuitive. If we can avoid it by endorsing one of my proposed views, so much the better.

Either of my formulations also avoid the metaphysical unwieldiness of epiphenomenal qualia to which the qualia-concerned physicalist, near-enough, may be committed. If Kim accepts qualia, he still needs to provide some sort of relation which holds between given qualia and physical properties. Why not, as I suggest, endorse one well-founded causal

relation as existing between all mental properties, qualia included? This move is at least as parsimonious as the type-identity view regarding relations- all it requires is treating certain relations as causal rather than identity, or some sort of thin non-identity for the NRPist. And this move permits one to dissolve epiphenomenal qualia, not just worrying on their face but also irreconcilable with Alexanders Dictum, as properties without causal import.

3.9.3 Accommodating Emergence

I have left the most controversial topic for NRP to the last. However, I do not plan on being *too* controversial here; I do not intend to defend emergence as being a true or accurate thesis about reality. What I want to assert is that emergence, as I will define it, *could* turn out to feature in our best scientific theories, and in our "map" of reality, and our formulation of physical to higher-level relations within physicalism ought not to rule out this possibility by fiat, even if the existence of emergent properties itself undermines physicalism. There is a fine line that needs to be drawn here. I do not wish to say that the relation between physical and mental properties one adopts, for NRP, ought to simply admit emergence. This alone has been levelled as a serious objection to supervenience formulations, and to Pereboom's "wholly made up of" relation.

Emergence comes in many forms in the literature. "Weak" sorts of emergence are non-metaphysical; a property might be called weakly emergent if it simply epistemically unpredictable from the micro-physical properties underpinning it. There is no suggestion of any truly novel or irreducible property in this sort of emergence-talk. The sort of emergence we are interested in is not so ontologically innocent.

In Section 3.8, I offered a definition of strong emergence (henceforth, simply emergence) as the view that 'mental properties have causal powers which are not explicable in terms of the causal powers of their physical substrates' (Crane, 2001, p.210-11). This is often characterised as a commitment to emergent properties having downward causal powers, such that a mental property could have some effect on lower-level physical properties which is not explicable in terms of the causal powers of its micro-physical "base"[100].

Of note here, emergence in general presents no problem for physicalism; only emergence at levels beyond the micro-physical. I mention this in light of an argument due to Campbell & Pickhard regarding configurational properties at the level of microphysics (2011, p.49).

As they note, emergent properties are novel, with downward causal powers not explicable by reference to their physical "base" or substrate. Here, they understand emergence as I have already described it. Further, they observe that 'particles participate in configurations relative to each other' (p.47). Quantum field theory reveals that the base level of reality may not be best understood as involving causation between micro-physical particulars;

'What our best contemporary physics reveals is that there are no elementary 'particles', elemental events, or some such particulars; everything is composed of quantum fields, of various [complexity] Quantum field theory shifts the basic ontology of the universe from micro-particles to quantum-fields-in-process. What have seemed to be 'particles' are now

[100] Robb (2018, p.165) presents a model of emergence such that it is empirically "invisible" in cases of mental causation. This is a non-standard notion of emergence, evidently not open to scientific investigation. I note this parenthetically, since I do not take it to pose any particular problems to my position.

conceptualized as particle-like processes and interactions resulting from the quantization of field processes and interactions.' (p.45)

Given this, it seems that our micro-physics contains emergent properties when understood as 'configurations of constituents [generating] the emergence of higher-level causal powers'(p.38). The causal powers of fields are not, seemingly, explicable merely by reference to the causal powers of their constituents. It may not even make scientific sense to *refer* to particular constituents or properties within these fields (p.49).

Campbell & Pickhard are not alone in postulating emergence within physics. Elsewhere, Schaffer makes a quite different argument for potential quantum properties being emergent (2010, p.50-3).

I will not concern myself here with concerns regarding the status of our base microphysical properties, whether fields are best understood within a process ontology or not (Campbell & Pickhard, 2011, p.45-7). Regardless, it does seem like the physicalist ought to be *willing* to admit the existence of configurational emergent properties at the level of physics. One can do this, it seems, without making any commitment to emergent properties at higher levels. Certainly, we do not *seem* to observe such properties at the level of psychology. This gives us reason, though, to seek an ontology of relations which can admit of emergent property relations as parsimoniously as possible. For reasons I now move to elaborate, I take it that the ontological commitments of W-grounding and the Causal View are fit to the task. My thinking runs as follows.

If emergence relations are discovered anywhere in the natural world, we should want that to undermine *NRP*, whatever formulation we adopt, but we should not want it to problematise the *relation* between mental and physical properties contained within that formulation. Further on that, we should not want it to be the case that our metaphysical picture, including our postulated relation between mental and physical properties, would have to be drastically amended to account for emergent properties or relations.

I have two reasons for thinking this. Physicalism is commonly understood as deriving from, or being motivated by, scientific discoveries (Papineau, 2000). As such, in that naturalistic spirit, physicalists ought to be open to the possibility of discoveries which undermine their adopted view here. However, the relations which we employ to capture inter-level dependencies fits into our ontology. While our ontology should not propose anything counter to empirical findings, it does not seem to me that it ought to be hostage to a particularised view of inter-level relations as we understand them today.

If supervenience physicalism permits emergence, then the supervenience relation is inappropriate. This is because, however, cases of emergence in this instance would not undermine the truth of that *formulation.* Emergence relations would be supervenient just like realisation-relations, for instance. On the other hand, W-grounding picks out a specific non-emergence relation. If it transpired, for instance, that phenomenal pain properties *specifically* were emergent, exerting downward causal power on later physical properties, this would constitute evidence against W-grounding physicalism. Since W-grounding would be the relation between any given mental and physical relata, an emergence relation

between some given mental-physical relata would be bad news for the view[101]. However, it would not mean that W-grounding itself could not adequately describe other, non-emergent mental-physical property relations.

Further, if emergent relations do turn out to be part of the fabric of reality, W-grounding and the Causal View are relationally well-equipped to accommodate them, albeit not within physicalist frameworks at that point. This is because both of my proposed views and emergentism understand mental properties as *caused* (Kim, 1999, p.32, O'Connor & Wong, 2005, p.671) and at a less fundamental level than their relevant causal bases. As such, emergent properties represent another caused, thickly non-identical higher-level non-fundamental property[102], not a major addition to our ontology.

Out of all this, we have two thick non-identity formulations which can both satisfy the physicalist commitments, albeit loosening NOAA to Nothing-Left, and which maintain the NRP theses of Irreducibility and Dependence. Besides this, they leave us well-disposed to accommodate irreducible qualia without resorting to dualism, alongside the other benefits just canvassed. They also avoid endorsement of *mere* nomological supervenience which marks out naturalistic dualism.

3.10 Concluding Thoughts; Where Do We Go From Here?

This paper listed three propositions at the outset, namely;

[101] Overall, of course. I take this very fact as good news for such a formulation insofar as it does not readily permit that to which physicalism is opposed.
[102] Some (Morgan, 1923) deny that emergence relations are causal. If this is the case, I must concede that this theoretical-parsimony advantage would not hold.

(1) Current formulations of non-reductive physicalism fail, or are liable to fail.

(2) Their failure stems from a common cause, specifically the sort of multiple realisability concern which motivates many to first adopt these positions.

(3) Loosening one commitment of non-reductive physicalism allows us to develop a more promising alternative formulation of the view.

After outlining the theses of NRP, and the commitments of physicalism more generally, I moved on to establishing (1). Beginning with token-identity, I argued that on a plausible property-exemplification view of events, token-identities threatened to collapse into type-identities. In addition, I presented two variations on the classic MR argument which demonstrated that token-identities would fail to hold for much the same reason as type-identities. I then considered standard realisation, as well as Shoemaker and Wilson's Subset Realisation view. This formulation draws an identity relation between causal powers, and hence falls prey to MR. My own MR arguments here rode alongside Pereboom's, though I dissented on his favoured formulation, a constitution physicalism which employs the "wholly made up of" relation. I argued that this relation was presented as basic, but that this basic nature was not intuitive in itself; rather it garnered its intuitive "oomph" from the identity relation which Pereboom intended to avoid. Further to this, I considered Lewis' likeness claims for composition and identity, noting that Pereboom's relation was substantially compatible with those claims, and that stipulations of directionality do little to assuage these concerns and save his relation as a throughgoing primitive in its own right. Then, I briefly glossed composition views, over which I take to hang the same sort of identity issue.

This left us in a position to wonder why these views might be going wrong, in ways which predictably tied back to identity and consequent MR concerns, bringing us to (2).

I suggest these formulations of NRP all demonstrate a commitment to thin non-identity, motivated by standard-NOAA. They want to avoid any ontologically immodest commitments to properties "over and above" the physical, and aim to get at this by developing mental-physical relations as close to identity as possible. This means that whatever relation they propose for conjoining mental and physical properties, it will tend to either involve some identity, whether between tokens or causal powers, vulnerable to MR objections, or it will invite worries that the relation is merely identity "in disguise". A thinly non-identical formulation which does not fail into this systematic trap may be possible, I concede. However, I suggest that the NRPist will have an easier time if they stop worrying and accept thick non-identity, which can be had for relatively little cost, and which carries its own theoretical benefits.

Regarding (3), I proposed that a loosening of Standard-NOAA, to Nothing-Left, allows us to accept thick non-identity. I hoped to demonstrate that formulations of NRP which endorse thick non-identity are viable insofar as they adhere to the core theses of NRP, Irreducibility and Dependence, as well as C2, and that they can broadly fit within the intuitive picture of physicalism in part captured by Nothing-Left. To recall Loewer, physicalism can be understood as the thesis that 'all God had to do to create our world was to create its physical facts and laws; the rest followed from these' (2009, p.39). Despite weakening NOAA to allow something "over" the physical, while wholly dependent upon it, and necessitated

causally or W-grounded in it, I think we preserve the vital parts of the physicalist picture, including the primacy of the physical and the all-else-generating status of physical properties captured in Loewer's thesis.

I considered several objections here. First, to deal with the worry that a more substantial distinctness of properties entailed by thick-non-identity leads us to naturalistic dualism, I demonstrate that Nothing-Left conjoined with Dependence suffices to rule out NRP-formulations which allow for metaphysically possible worlds where the mental does not supervene on the physical. Further, I outline two formulations which make good on this; W-grounding and my Causal View. Both posit causal, necessitating, and importantly *dependent* relations between physical and mental properties.

Pre-empting concerns about ontological innocence, while I concede that these thick-NRP formulations present additions to being, I argued that they provide us with a *more* parsimonious view of inter-property relations since both are fundamentally causal notions. Causation is already part of the ontology of physicalists I am concerned with, so they need only refine that notion, and they can perhaps rid themselves of reliance on thin non-identity relations. Beyond this, I think we can reasonably assert that both thick-NRP formulations do a better job of satisfying the core of our non-reductive intuitions, when it comes to Irreducibility of the mental. We are presenting throughgoing non-identity by way of genuinely distinct mental properties, nevertheless wholly dependent upon physical properties. At least, I think this should give us some hope that endorsing thick non-identity is an option going forward for NRP, in light of the failures of thin-NRP formulations.

Finally, I considered three philosophical benefits of adopting thick-NRP, in either formulation. They provide us, alongside other argumentation, with a convincing causal story regarding mental causation. They also allow us to dissolve worries about qualia and phenomenal properties quite generally, while avoiding naturalistic dualism. Further, they provide relations which, while themselves precluding strong emergence, could nevertheless continue to feature comfortably and without significant ontological addition, in the world where physicalism itself was undermined and emergent properties did happen to exist at some level of reality above the microphysical[103].

Perhaps W-grounding is open to some objections in principle that have not been considered here. Further, my Causal View's synchronic causation may have some caveats one is inclined to regard as unacceptable. Nevertheless, if I have demonstrated anything, I hope it is that NRP can survive with thick non-identity. The challenge for those truly concerned with MR, and likewise concerned with present formulations of NRP, is to develop a thick non-identity relation which can be applied to the relation between the mental and the physical, while preserving the intuitive physicalist picture and relevant commitments. If we accept thick non-identity, the space for formulating NRP expands drastically. As I see it, the central takeaway from all of this is that, so long as we satisfy the substantive commitments of NRP and physicalism, while maintaining the intuitive physicalist picture, we can afford to be more expansive in how we formulate our best model of NRP.

[103] Which is to say, physicalists can adopt a framework which, while physicalist, need not be thrown out "baby and bathwater" in light of hypothetical future scientific discoveries which would undercut physicalism itself.

Bibliography

Adlam, E. (2022) 'Two Roads to Retrocausality', *Synthese,* 200(5), pp. 1-36.

Árnadóttir, S.T. & Crane, T. (2013) 'There is No Exclusion Problem' in Lowe, E.J, Gibb, S.C. & Ingthorsson, R.D. (eds.), *Mental Causation and Ontology*, Oxford: Oxford University Press, pp. 248-66.

Armstrong, D.M. (1986) 'In Defence of Structural Universals', *Australasian Journal of Philosophy,* 64 (1), pp. 85–88.

Armstrong, D.M. (1997) *A World of States of Affairs*, Cambridge: Cambridge University Press.

Baker, A. (2016) 'Parsimony and inference to the best mathematical explanation', *Synthese*, 193 (2), pp. 333-350.

Bartels, A. & Wohlfarth, D. (2015) 'How Fundamental Physics Represents Causality' in *Recent Developments in the Philosophy of Science: EPSA13 Helsinki*, pp.197-207.

Baxter, D. (2001) 'Instantiation as Partial Identity', *The Australasian Journal of Philosophy*, 79(4), pp. 449–464.

Baysan, U. (2020) 'Causal Emergence and Epiphenomenal Emergence', *Erkenntnis,* 85(4), pp. 891-904.

Bernstein, S. (2016) 'Overdetermination underdetermined', *Erkenntnis*, 81, pp.17-40.

Bishop, R.C. (2006) 'The hidden premise in the causal argument for physicalism', *Analysis*, 66(1), pp. 44-52

Block, N. (2003) 'Do causal powers drain away?', *Philosophy and Phenomenological Research*, 67(1), pp. 133-150.

Block, N.J. & Jerry A.F. (1972) ‘What Psychological States Are Not’, *The Philosophical Review*, 81(2), pp. 159–181.

Bonjour, L. (2009) ‘Against Materialism’, in Koons, R. & Bealer. G. (eds.), *The Waning of Materialism*. Oxford: Oxford University Press.

Bontley, T.D. (2005) ‘Proportionality, causation, and exclusion’, *Philosophia*, 32(1-4), pp. 331-348.

Boyd, R., Gasper, P. & Trout, J.D. (1991) *The philosophy of science*, Cambridge MA: MIT Press.

Burge, T. (2007) *Foundations of mind*, New York: Oxford University Press.

Campbell, R.J. & Bickhard, M.H. (2011) ‘Physicalism, emergence and downward causation’, *Axiomathes*, 21, pp. 33-56.

Campbell, N. (2015) ‘Does Same-Level Causation Entail Downward Causation?’ *Abstracta*, 8(2), pp. 53-66.

Cargile, J. (2003) ‘On "Alexander's" dictum’, *Topoi*, 22(2). pp. 143-149.

Chalmers, D.J. (1996) *The conscious mind: In search of a fundamental theory*, Oxford: Oxford University Press.

Chalmers, D. (2004) ‘The Representational Character of Experience’ in Leiter, B. (ed.) *The Future for Philosophy*, Oxford: Oxford University Press.

Charles, D. (1992) ‘Supervenience, composition, and physicalism’, in Charles, D. & Lennon, K. (eds.), *Reduction, Explanation and Realism,* Oxford: Oxford University Press.

Crane, T. (1995) ‘The Mental Causation Debate’, *Proceedings of the Aristotelian Society*, Vol. LXIX.

Crane, T. (2001) 'The Significance of Emergence', in Loewer, B. & Gillett, G. (eds.), *Physicalism and its Discontents*, Cambridge, UK: Cambridge University Press. pp. 207-224.

Daly, C. (1998) 'What are physical properties?', *Pacific Philosophical Quarterly,* 79(3), pp. 196-217.

Davidson, D. (1970) '*Mental Events',* in Foster, L. & Swanson, J. W. (eds.), *Experience and Theory*, Oxford: Clarendon Press. pp. 207-224.

Fodor, J. A. (1974) 'Special Sciences (or: The Disunity of Science as a Working Hypothesis)', *Synthese*, 28(2), pp. 97–115.

Foster, J. (1991) *The Immaterial Self: A Defence of the Cartesian Dualist Conception of the Mind*. London: Routledge.

Funkhouser, E. (2002) 'Three varieties of causal overdetermination', *Pacific philosophical quarterly*, 83(4), pp. 335-351.

Gibb, S. (2010) 'Closure principles and the laws of conservation of energy and momentum', *Dialectica*, *64*(3), pp. 363-384.

Gibb, S. (2015.a) 'The Causal Closure Principle', *The Philosophical Quarterly,* 65(261), pp. 626-647

Gibb, S. (2015.b) 'Defending Dualism', *Proceedings of the Aristotelian Society,* 115, pp. 131-146.

Gibb, S. (2015.c) 'Physical determinability', *Humana mente: Journal of Philosophical Studies*, 8(29), pp. 69-90.

Gibbons, J. (2006) ‘Mental causation without downward causation’, *Philosophical Review*, 115(1), pp. 79-103.

Gillett, C. (2002) ‘Strong Emergence as a Defense of Non-Reductive Physicalism’, *Principia*, 6(1), pp. 87-120.

Goff, P. (2016) ‘Fundamentality and the Mind-Body Problem’, *Erkenntnis*, 81(4), pp. 881-898.

Goff, P. (2017) *Consciousness and Fundamental Reality*, New York: Oxford University Press.

Grygianiec, M. (2016) ‘How to Get Rid of Closure’, *Diametros*, 48, pp. 1-17.

Hall, N. (2004) ‘Two concepts of causation’, in Collins, J., Hall, N. & Paul, L. (eds.), *Causation and counterfactuals*, , Cambridge MA: MIT Press, pp. 225-276.

Hart, W. D. (1988) *The Engines of the Soul*, New York: Cambridge University Press.

Hawley, K. (2014) ‘Ontological Innocence’, in Cotnoir, A. J. & Baxter, D. L. M (eds.), *Composition as Identity*, Oxford: Oxford University Press. pp. 70-89.

Hodes, H. T. (1990) ‘Ontological Commitments, Thick and Thin’, in Boolos, G. (ed.), *Method, Reason and Language: Essays in Honor of Hilary Putnam,* Cambridge UK: Cambridge University Press. pp. 235-260.

Horgan, T. (1984) Functionalism and token physicalism, *Synthese*, 59(3), pp. 321-338.

Jackson, F. (1982) ‘Epiphenomenal qualia’, *The Philosophical Quarterly*, 32(127), pp.127-136.

Kallestrup, J. (2006) ‘The causal exclusion argument’, *Philosophical Studies*, 131, pp. 459-485.

Kim, J (1992) ‘Multiple realization and the metaphysics of reduction’, *Philosophy and Phenomenological Research,* 52(1), pp. 1-26.

Kim, J. (1993), 'Mental causation in a physical world', *Philosophical issues*, 3, pp.157-176.

Kim, J. (1999) 'Making sense of emergence', *Philosophical Studies* 95(1-2), pp. 3-36.

Kim, J. (2003) 'Blocking Causal Drainage and Other Maintenance Chores with Mental Causation', *Philosophy and Phenomenological Research*, 67(1), pp. 151-176.

Kim, J. (2005) *Physicalism, or something near enough,* Princeton: Princeton University Press.

Kim, J. (2007) 'Causation and mental causation' in McLaughlin, B. & Cohen, J. (eds.), *Contemporary Debates in Philosophy of Mind,* Oxford: Blackwell, pp. 227-242.

Kim, J. (2012) 'The very idea of token physicalism' in Hill, C. & Gozzano, S. (eds.), *New Perspectives on Type Identity: The Mental and the Physical*, Cambridge UK: Cambridge University Press, p. 167.

Lebender, D. and Schneider, F.W. (1994) 'Logical gates using a nonlinear chemical reaction', *The Journal of Physical Chemistry*, *98*(31), pp.7533-7537.

Langton, R. & Lewis, D. (1998) 'Defining "intrinsic"', *Philosophy and Phenomenological Research*, 58(2), pp. 333-345.

LePore, E. & Loewer, B. (1989) 'More on Making Mind Matter', *Philosophical Topics,* 17(1), pp. 175-191.

Lewis, D. (1973) *Counterfactuals*, Malden, MA: Blackwell.

Lewis, D. (1983) *Philosophical papers,* New York: Oxford University Press.

Lewis, D. (1991) *Parts of Classes*, Oxford: Basil Blackwell

Lim, D (2014) 'Occasionalism and non-reductive physicalism: another look at the continuous creation argument', *International Journal for Philosophy of Religion,* 75(1), pp. 39-57.

List, C. (2019) ‘Levels: descriptive, explanatory, and ontological’, *Noûs*, 53(4), pp. 852-883.

List, C. & Menzies, P. (2009) ‘Nonreductive Physicalism and the Limits of the Exclusion Principle’, *The Journal of Philosophy*, 106(9), pp. 475-502.

Loew, C. (2017), ‘Causation, physics, and fit’, *Synthese,* 194(6), pp. 1945–1965.

Loewer, B. & Gillett, G. (eds.) (2001) *Physicalism and its Discontents*, Cambridge: Cambridge University Press.

Loewer, B. (2007) ‘Mental causation, or something near enough’, *Contemporary debates in philosophy of mind*, pp. 243-264.

Lowe, E.J. (2000) ‘Causal closure principles and emergentism’, *Philosophy*, *75*(04).

Lowe, E. J. (2009) ‘Substance dualism: A non-cartesian approach’, in Koons, R.C. & Bealer, G. (eds.), *The Waning of Materialism: New Essays*, Oxford: Oxford University Press, pp. 439-462.

Mackie, J.L. (1965) ‘Causes and Conditions’, *American Philosophical Quarterly*, 2(4), pp. 245–264.

Marras, A. (2007) ‘Kim’s Supervenience Argument and Nonreductive Physicalism’, *Erkenntnis*, 66(3), pp. 305 - 327.

McLaughlin, B.P. (2007), ‘Mental causation and Shoemaker-realization’, *Erkenntnis*, 67(2), pp. 149-172.

Meixner, U. (2008) ‘New Perspective for a Dualistic Conception of Mental Causation’, *Journal of Consciousness Studies,* 15(1), pp. 17-38.

Mellor, D.H. (1995) *The Facts of Causation*, New York: Routledge.

Melnyk, A. (1994) 'Being a physicalist: How and (more importantly) why', *Philosophical Studies,* 74(2), pp. 221-241.

Melnyk, A. (1995) 'Two cheers for reductionism, or, the dim prospects for nonreductive materialism', *Philosophy of Science,* 62(3), pp. 370-88.

Melnyk, A. (2003) 'Some Evidence for Physicalism', in Walter, S. & Heckmann, H. (eds.), *Physicalism and Mental Causation*, Exeter: Imprint Academic, pp. 155-172.

Melnyk, A. (2006) 'Realization and the Formulation of Physicalism', *Philosophical Studies*, 131(1), pp. 127–155.

Melnyk, A. (2008) 'Can Physicalism Be Non-Reductive?', *Philosophy Compass*, 3(6), pp. 1281-1296.

Menzies, P. (2007) 'Causation in context. Causation, physics, and the constitution of reality', pp.191-223.

Menzies, P. (2013) 'Mental causation in the physical world', in Gibb, S.C., Lowe, E.J. & Ingthorsson, R.D. (eds.), *Mental Causation and Ontology*, Oxford: Oxford University Press. pp. 58-87.

Merricks, T. (2001) *Objects and Persons,* New York: Oxford University Press.

Montero, B. (2003) 'Varieties of causal closure', in Sven, W. (ed.) *Physicalism and Mental Causation: the metaphysics of mind and action*, Exeter: Imprint Academic, pp. 173-187.

Moore, G.E. (1922) 'The Conception of Intrinsic Value', in Rachels, J. (ed.) (1998) *Philosophical Studies.* Oxford: Oxford University Press.

Moore, D (2012) ‘Causal Exclusion and Dependent Overdetermination’, *Erkenntnis,* 76(3), pp. 319-335.

Morgan. C.L. (1923), ‘Emergence’, in *Emergent Evolution*, London: Williams and Norgate, pp. 1-34.

Nagel, T. (1986) *The View From Nowhere*. New York: Oxford University Press.

Ney, A. (2016) ‘Microphysical Causation and the Case for Physicalism’, *Analytic Philosophy,* 57(1), pp. 141-164.

Noordhof, P. (2013) ‘Mental Causation: Ontology and Patterns of Variation, in Gibb, S.C., Lowe, E.J. & Ingthorsson, R.D. (eds.), *Mental Causation and Ontology*, Oxford: Oxford University Press. pp. 88-125.

O’Connor, T. & Wong, H.Y. (2005) ‘The Metaphysics of Emergence’, *Noûs*, 39(4), pp. 658–678.

Papineau, D. (1991) ‘The Reason Why: Response to Crane’, *Analysis*, 51(1), pp. 37-40.

Papineau, D. (2001) ‘The Rise of Physicalism’, in Gillett. C. & Loewer, B.M. (eds.) *Physicalism and its Discontents*, Cambridge: Cambridge University Press.

Papineau, D. (2013) ‘Causation is macroscopic but not irreducible’ in Gibb, S.C., Lowe, E.J. & Ingthorsson, R.D. (eds.), *Mental Causation and Ontology*, Oxford: Oxford University Press. p. 126.

Paul, L.A. & Hall, N. (2013) *Causation: A User’s Guide*, Oxford: Oxford University Press.

Peacocke, C. (1977) *Holistic Explanation*, Oxford: Clarendon Press.

Pereboom, D. (2002) ‘On Baker’s Persons and Bodies’, *Philosophy and Phenomenological Research,* 64(3), pp.615–622.

Pereboom, D. (2011) *Consciousness and the Prospects of Physicalism*, New York: Oxford University Press.

Pereboom, D. & Kornblith, H. (2003), 'The Metaphysics of Irreducibility', in Heil, J. (ed.), *Philosophy of Mind: A Guide and Anthology*, Oxford: Oxford University Press.

Pettit, P. (1993) 'A definition of physicalism', *Analysis*, *53*(4), pp. 213-223.

Quine, W.V.O. (1981) *Theories and things*, Harvard: Harvard University Press.

Quine, W.V.O. (1984) 'Relativism and absolutism', *The Monist*, 67(3), pp. 293-296.

Raatikainen, P. (2018) 'Kim on Causation and Mental Causation', *E-Logos Electronic Journal for Philosophy,* 25(2), pp. 22–47.

Robb, D. (2018) 'Could Mental Causation Be Invisible?' in Carruth, A., Gibb S.C. & Heil, J. (eds.), *The Metaphysics of E.J. Lowe*, Oxford: Oxford University Press.

Rosen, G. (2009) 'Metaphysical dependence: Grounding and reduction', in Hale, B. & Hoffman, A. (eds.), *Modality: Metaphysics, Logic, and Epistemology*, pp. 109-135.

Rueger, A. (2000) 'Physical emergence, diachronic and synchronic', *Synthese,* 124(3), pp. 297-322.

Schaffer, J. (2003) 'Overdetermining causes', *Philosophical Studies,* 114(1-2), pp. 23 - 45.

Schaffer, J. (2010) 'Monism: The Priority of the Whole', *Philosophical Review,* 119(1), pp. 31-76.

Schneider, S. (2013) 'Non-reductive physicalism and the mind problem', *Noûs*, 47(1), pp. 135-153.

Schwarz, W. (2013) 'Contingent identity', *Philosophy Compass*, 8(5), pp. 486-495.

Schwichtenberg, J. (2018) *Physics from symmetry,* New York: Springer.

Shapiro, L.A. (2002) *Neural plasticity and multiple realizability*

Shoemaker, S. (2007) *Physical realization,* Oxford: Clarendon Press.

Shoemaker, S. (2013) 'Physical realization without pre-emption', in Gibb, S.C & Ingthorsson, R. (eds.), *Mental Causation and Ontology,* Oxford: Oxford University Press.

Sider, T. (2003) 'What's so bad about overdetermination?', *Philosophical and Phenomenological Research*, 67(3), pp. 719-726.

Skow, B (2013) 'Are There Non-Causal Explanations (of Particular Events)?', *British Journal for the Philosophy of Science,* (3)47.

Smart, J.J. (1959) 'Sensations and brain processes', *The Philosophical Review*, 68(2), pp. 141-156.

Spurrett, D. & Papineau, D. (1999) 'A Note on the Completeness of "Physics"', *Analysis*, *59*(1), pp. 25-29.

Stenwall, R. (2020) 'A grounding physicalist solution to the causal exclusion problem', *Synthese,* 198(12), pp. 11775-11795.

Stoljar, D. (2010) *Physicalism*, New York: Routledge

Strawson, G. (2006) 'Realistic monism: why physicalism entails panpsychism', in Freeman, A. (ed.), *Consciousness and its place in nature: does physicalism entail panpsychism?*, Exeter: Imprint Academic, pp. 3-31.

Sturgeon, S. (1998) 'Physicalism and overdetermination', *Mind,* 107(426), pp. 411-432.

Sturgeon, S. (2000) *Matters of Mind: Consciousness, Reason and Nature*, London: Psychology Press.

Trogdon, K. (2009) 'Physicalism and sparse ontology', *Philosophical Studies,* 143(2), pp. 147-165.

Van Inwagen, P. (1994) 'Composition as Identity', *Philosophical Perspectives,* 8, pp. 207 - 220.

Vicente, A. (2006) 'On the causal completeness of physics', *International Studies in the Philosophy of Science*, 20(2), pp. 149-171.

Hansson Wahlberg, T. (2022) 'Sparse Causation and Mere Abundant Causation', *Philosophical Studies,* 179(11), pp. 3259-3280.

Wilson, A. (2018) 'Metaphysical Causation', *Noûs,* 520(4), pp. 723-751.

Yablo, S. (1992) 'Mental causation', *The Philosophical Review*, *101*(2), pp. 245-280.

Zangwill, N. (2005) 'The normativity of the mental', *Philosophical Explorations*, 8(1), pp. 1-19.